Chapter 7

PARAMETER ESTIMATION

7.1 INTRODUCTION

Let $X_1, \ldots, X_n$ be a random sample from a distribution F_θ that is specified up to a vector of unknown parameters θ. For instance, the sample could be from a Poisson distribution whose mean value is unknown; or it could be from a normal distribution having an unknown mean and variance. Whereas in probability theory it is usual to suppose that all of the parameters of a distribution are known, the opposite is true in statistics, where a central problem is to use the observed data to make inferences about the unknown parameters.

In Section 7.2, we present the *maximum likelihood* method for determining estimators of unknown parameters. The estimates so obtained are called *point estimates*, because they specify a single quantity as an estimate of θ. In Section 7.3, we consider the problem of obtaining *interval estimates*. In this case, rather than specifying a certain value as our estimate of θ, we specify an interval in which we estimate that θ lies. Additionally, we consider the question of how much *confidence* we can attach to such an interval estimate. We illustrate by showing how to obtain an interval estimate of the unknown mean of a normal distribution whose variance is specified. We then consider a variety of interval estimation problems. In Section 7.3.1, we present an interval estimate of the mean of a normal distribution whose variance is unknown. In Section 7.3.2, we obtain an interval estimate of the variance of a normal distribution. In Section 7.4, we determine an interval estimate for the difference of two normal means, both when their variances are assumed to be known and when they are assumed to be unknown (although in the latter case we suppose that the unknown variances are equal). In Sections 7.5 and the optional Section 7.6, we present interval estimates of the mean of a Bernoulli random variable and the mean of an exponential random variable.

In the optional Section 7.7, we return to the general problem of obtaining point estimates of unknown parameters and show how to evaluate an estimator by considering its mean square error. The bias of an estimator is discussed, and its relationship to the mean square error is explored.

In the optional Section 7.8, we consider the problem of determining an estimate of an unknown parameter when there is some prior information available. This is the *Bayesian* approach, which supposes that prior to observing the data, information about θ is always available to the decision maker, and that this information can be expressed in terms of a probability distribution on θ. In such a situation, we show how to compute the *Bayes estimator*, which is the estimator whose expected squared distance from θ is minimal.

7.2 MAXIMUM LIKELIHOOD ESTIMATORS

Any statistic used to estimate the value of an unknown parameter θ is called an *estimator* of θ. The observed value of the estimator is called the *estimate*. For instance, as we shall see, the usual estimator of the mean of a normal population, based on a sample $X_1, \dots, X_n$ from that population, is the sample mean $\overline{X} = \sum_i X_i/n$. If a sample of size 3 yields the data $X_1 = 2, X_2 = 3, X_3 = 4$, then the estimate of the population mean, resulting from the estimator $\overline{X}$, is the value 3.

Suppose that the random variables $X_1, \dots, X_n$, whose joint distribution is assumed given except for an unknown parameter θ, are to be observed. The problem of interest is to use the observed values to estimate θ. For example, the X_i's might be independent, exponential random variables each having the same unknown mean θ. In this case, the joint density function of the random variables would be given by

$$
\begin{aligned}
&f(x_1, x_2, \dots, x_n)\\
&\quad = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n)\\
&\quad = \frac{1}{\theta} e^{-x_1/\theta} \frac{1}{\theta} e^{-x_2/\theta} \cdots \frac{1}{\theta} e^{-x_n/\theta}, \qquad 0 < x_i < \infty, i = 1, \dots, n\\
&\quad = \frac{1}{\theta^n} \exp\left\{-\sum_1^n x_i/\theta\right\}, \qquad 0 < x_i < \infty, i = 1, \dots, n
\end{aligned}
$$

and the objective would be to estimate θ from the observed data $X_1, X_2, \dots, X_n$.

A particular type of estimator, known as the *maximum likelihood* estimator, is widely used in statistics. It is obtained by reasoning as follows. Let $f(x_1, \dots, x_n|\theta)$ denote the joint probability mass function of the random variables $X_1, X_2, \dots, X_n$ when they are discrete, and let it be their joint probability density function when they are jointly continuous random variables. Because θ is assumed unknown, we also write f as a function of θ. Now since $f(x_1, \dots, x_n|\theta)$ represents the likelihood that the values $x_1, x_2, \dots, x_n$ will be observed when θ is the true value of the parameter, it would seem that a reasonable estimate of θ would be that value yielding the largest likelihood of the observed values. In other words, the maximum likelihood estimate $\hat{\theta}$ is defined to be that value of θ maximizing $f(x_1, \dots, x_n|\theta)$ where $x_1, \dots, x_n$ are the observed values. The function $f(x_1, \dots, x_n|\theta)$ is often referred to as the *likelihood* function of θ.

In determining the maximizing value of θ, it is often useful to use the fact that $f(x_1, \ldots, x_n|\theta)$ and $\log[f(x_1, \ldots, x_n|\theta)]$ have their maximum at the same value of θ. Hence, we may also obtain $\hat{\theta}$ by maximizing $\log[f(x_1, \ldots, x_n|\theta)]$.

EXAMPLE 7.2a (Maximum Likelihood Estimator of a Bernoulli Parameter) Suppose that n independent trials, each of which is a success with probability p, are performed. What is the maximum likelihood estimator of p?

SOLUTION The data consist of the values of $X_1, \ldots, X_n$ where

$$X_i = \begin{cases} 1 & \text{if trial } i \text{ is a success} \\ 0 & \text{otherwise} \end{cases}$$

Now

$$P\{X_i = 1\} = p = 1 - P\{X_i = 0\}$$

which can be succinctly expressed as

$$P\{X_i = x\} = p^x(1-p)^{1-x}, \quad x = 0, 1$$

Hence, by the assumed independence of the trials, the likelihood (that is, the joint probability mass function) of the data is given by

$$\begin{aligned} f(x_1, \ldots, x_n|p) &= P\{X_1 = x_1, \ldots, X_n = x_n|p\} \\ &= p^{x_1}(1-p)^{1-x_1} \cdots p^{x_n}(1-p)^{1-x_n} \\ &= p^{\sum_1^n x_i}(1-p)^{n-\sum_1^n x_i}, \quad x_i = 0, 1, \quad i = 1, \ldots, n \end{aligned}$$

To determine the value of p that maximizes the likelihood, first take logs to obtain

$$\log f(x_1, \ldots, x_n|p) = \sum_1^n x_i \log p + \left(n - \sum_1^n x_i\right) \log(1-p)$$

Differentiation yields

$$\frac{d}{dp} \log f(x_1, \ldots, x_n|p) = \frac{\sum_1^n x_i}{p} - \frac{\left(n - \sum_1^n x_i\right)}{1-p}$$

Upon equating to zero and solving, we obtain that the maximum likelihood estimate $\hat{p}$ satisfies

$$\frac{\sum_{1}^{n} x_i}{\hat{p}} = \frac{n - \sum_{1}^{n} x_i}{1 - \hat{p}}$$

or

$$\hat{p} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Hence, the maximum likelihood estimator of the unknown mean of a Bernoulli distribution is given by

$$d(X_1, \ldots, X_n) = \frac{\sum_{i=1}^{n} X_i}{n}$$

Since $\sum_{i=1}^{n} X_i$ is the number of successful trials, we see that the maximum likelihood estimator of p is equal to the proportion of the observed trials that result in successes. For an illustration, suppose that each RAM (random access memory) chip produced by a certain manufacturer is, independently, of acceptable quality with probability p. Then if out of a sample of 1,000 tested 921 are acceptable, it follows that the maximum likelihood estimate of p is .921. ■

EXAMPLE 7.2b Two proofreaders were given the same manuscript to read. If proofreader 1 found n_1 errors, and proofreader 2 found n_2 errors, with $n_{1,2}$ of these errors being found by both proofreaders, estimate N, the total number of errors that are in the manuscript.

SOLUTION Before we can estimate N we need to make some assumptions about the underlying probability model. So let us assume that the results of the proofreaders are independent, and that each error in the manuscript is independently found by proofreader i with probability p_i, $i = 1, 2$.

To estimate N, we will start by deriving an estimator of p_1. To do so, note that each of the n_2 errors found by reader 2 will, independently, be found by proofreader 1 with probability p_i. Because proofreader 1 found $n_{1,2}$ of those n_2 errors, a reasonable estimate of p_1 is given by

$$\hat{p}_1 = \frac{n_{1,2}}{n_2}$$

However, because proofreader 1 found n_1 of the N errors in the manuscript, it is reasonable to suppose that p_1 is also approximately equal to $\frac{n_1}{N}$. Equating this to $\hat{p}_1$ gives that

$$\frac{n_{1,2}}{n_2} \approx \frac{n_1}{N}$$

or

$$N \approx \frac{n_1 n_2}{n_{1,2}}$$

Because the preceding estimate is symmetric in n_1 and n_2, it follows that it is the same no matter which proofreader is designated as proofreader 1.

An interesting application of the preceding occurred when two teams of researchers recently announced that they had decoded the human genetic code sequence. As part of their work both teams estimated that the human genome consisted of approximately 33,000 genes. Because both teams independently arrived at the same number, many scientists found this number believable. However, most scientists were quite surprised by this relatively small number of genes; by comparison it is only about twice as many as a fruit fly has. However, a closer inspection of the findings indicated that the two groups only agreed on the existence of about 17,000 genes. (That is, 17,000 genes were found by both teams.) Thus, based on our preceding estimator, we would estimate that the actual number of genes, rather than being 33,000, is

$$\frac{n_1 n_2}{n_{1,2}} = \frac{33{,}000 \times 33{,}000}{17{,}000} \approx 64{,}000$$

(Because there is some controversy about whether some of genes claimed to be found are actually genes, 64,000 should probably be taken as an upper bound on the actual number of genes.)

The estimation approach used when there are two proofreaders does not work when there are m proofreaders, when $m > 2$. For, if for each i, we let $\hat{p}_i$ be the fraction of the errors found by at least one of the other proofreaders j, $(j \neq i)$, that are also found by i, and then set that equal to $\frac{n_i}{N}$, then the estimate of N, namely $\frac{n_i}{\hat{p}_i}$, would differ for different values of i. Moreover, with this approach it is possible that we may have that $\hat{p}_i > \hat{p}_j$ even if proofreader i finds fewer errors than does proofreader j. For instance, for $m = 3$, suppose proofreaders 1 and 2 find exactly the same set of 10 errors whereas proofreader 3 finds 20 errors with only 1 of them in common with the set of errors found by the others. Then, because proofreader 1 (and 2) found 10 of the 29 errors found by at least one of the other proofreaders, $\hat{p}_i = 10/29$, $i = 1, 2$. On the other hand, because proofreader 3 only found 1 of the 10 errors found by the others, $\hat{p}_3 = 1/10$. Therefore, although proofreader 3 found twice the number of errors as did proofreader 1, the estimate of p_3 is less than that of p_1. To obtain more reasonable estimates, we could take the preceding values of $\hat{p}_i$, $i = 1, \ldots, m$, as preliminary estimates of the p_i. Now, let n_f be the number of errors that are found by at least one proofreader. Because n_f/N is the fraction of errors that are found by at least one proofreader, this should approximately equal $1 - \prod_{i=1}^{m}(1 - p_i)$, the probability that an error is found by at least one proofreader. Therefore, we have

$$\frac{n_f}{N} \approx 1 - \prod_{i=1}^{m}(1 - p_i)$$

suggesting that $N \approx \hat{N}$, where

$$\hat{N} = \frac{n_f}{1 - \prod_{i=1}^{m}(1 - \hat{p}_i)} \tag{7.2.1}$$

With this estimate of N, we can then reset our estimates of the p_i by using

$$\hat{p}_i = \frac{n_i}{\hat{N}}, \quad i = 1, \ldots, m \tag{7.2.2}$$

We can then reestimate N by using the new value (Equation 7.2.1). (The estimation need not stop here; each time we obtain a new estimate $\hat{N}$ of N we can use Equation 7.2.2 to obtain new estimates of the p_i, which can then be used to obtain a new estimate of N, and so on.) ■

EXAMPLE 7.2c (Maximum Likelihood Estimator of a Poisson Parameter) Suppose $X_1, \ldots, X_n$ are independent Poisson random variables each having mean λ. Determine the maximum likelihood estimator of λ.

SOLUTION The likelihood function is given by

$$\begin{aligned} f(x_1, \ldots, x_n|\lambda) &= \frac{e^{-\lambda}\lambda^{x_1}}{x_1!} \cdots \frac{e^{-\lambda}\lambda^{x_n}}{x_n!} \\ &= \frac{e^{-n\lambda}\lambda^{\Sigma_1^n x_i}}{x_1! \ldots x_n!} \end{aligned}$$

Thus,

$$\log f(x_1, \ldots, x_n|\lambda) = -n\lambda + \sum_{1}^{n} x_i \log \lambda - \log c$$

where $c = \prod_{i=1}^{n} x_i!$ does not depend on λ, and

$$\frac{d}{d\lambda} \log f(x_1, \ldots, x_n|\lambda) = -n + \frac{\sum_{1}^{n} x_i}{\lambda}$$

By equating to zero, we obtain that the maximum likelihood estimate $\hat{\lambda}$ equals

$$\hat{\lambda} = \frac{\sum_{1}^{n} x_i}{n}$$

and so the maximum likelihood estimator is given by

$$d(X_1, \ldots, X_n) = \frac{\sum_{i=1}^{n} X_i}{n}$$

For example, suppose that the number of people who enter a certain retail establishment in any day is a Poisson random variable having an unknown mean λ, which must be estimated. If after 20 days a total of 857 people have entered the establishment, then the maximum likelihood estimate of λ is $857/20 = 42.85$. That is, we estimate that on average, 42.85 customers will enter the establishment on a given day. ■

EXAMPLE 7.2d The number of traffic accidents in Berkeley, California, in 10 randomly chosen nonrainy days in 1998 is as follows:

$$4, 0, 6, 5, 2, 1, 2, 0, 4, 3$$

Use these data to estimate the proportion of nonrainy days that had 2 or fewer accidents that year.

SOLUTION Since there are a large number of drivers, each of whom has a small probability of being involved in an accident in a given day, it seems reasonable to assume that the daily number of traffic accidents is a Poisson random variable. Since

$$\overline{X} = \frac{1}{10} \sum_{i=1}^{10} X_i = 2.7$$

it follows that the maximum likelihood estimate of the Poisson mean is 2.7. Since the long-run proportion of nonrainy days that have 2 or fewer accidents is equal to $P\{X \leq 2\}$, where X is the random number of accidents in a day, it follows that the desired estimate is

$$e^{-2.7}(1 + 2.7 + (2.7)^2/2) = .4936$$

That is, we estimate that a little less than half of the nonrainy days had 2 or fewer accidents. ■

EXAMPLE 7.2e (Maximum Likelihood Estimator in a Normal Population) Suppose $X_1, \ldots, X_n$ are independent, normal random variables each with unknown mean μ and unknown standard deviation σ. The joint density is given by

$$f(x_1, \ldots, x_n | \mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{-(x_i - \mu)^2}{2\sigma^2}\right]$$

$$= \left(\frac{1}{2\pi}\right)^{n/2} \frac{1}{\sigma^n} \exp\left[\frac{-\sum_1^n (x_i - \mu)^2}{2\sigma^2}\right]$$

The logarithm of the likelihood is thus given by

$$\log f(x_1, \ldots, x_n | \mu, \sigma) = -\frac{n}{2}\log(2\pi) - n\log\sigma - \frac{\sum_1^n (x_i - \mu)^2}{2\sigma^2}$$

In order to find the value of μ and σ maximizing the foregoing, we compute

$$\frac{\partial}{\partial\mu} \log f(x_1, \ldots, x_n | \mu, \sigma) = \frac{\sum_{i=1}^n (x_i - \mu)}{\sigma^2}$$

$$\frac{\partial}{\partial\sigma} \log f(x_1, \ldots, x_n | \mu, \sigma) = -\frac{n}{\sigma} + \frac{\sum_1^n (x_i - \mu)^2}{\sigma^3}$$

Equating these equations to zero yields that

$$\hat{\mu} = \sum_{i=1}^n x_i/n$$

and

$$\hat{\sigma} = \left[\sum_{i=1}^n (x_i - \hat{\mu})^2/n\right]^{1/2}$$

Hence, the maximum likelihood estimators of μ and σ are given, respectively, by

$$\overline{X} \quad \text{and} \quad \left[\sum_{i=1}^{n}(X_i - \overline{X})^2/n\right]^{1/2} \tag{7.2.3}$$

It should be noted that the maximum likelihood estimator of the standard deviation σ differs from the sample standard deviation

$$S = \left[\sum_{i=1}^{n}(X_i - \overline{X})^2/(n-1)\right]^{1/2}$$

in that the denominator in Equation 7.2.3 is $\sqrt{n}$ rather than $\sqrt{n-1}$. However, for n of reasonable size, these two estimators of σ will be approximately equal. ■

EXAMPLE 7.2f *Kolmogorov's law of fragmentation* states that the size of an individual particle in a large collection of particles resulting from the fragmentation of a mineral compound will have an approximate lognormal distribution, where a random variable X is said to have a *lognormal* distribution if $\log(X)$ has a normal distribution. The law, which was first noted empirically and then later given a theoretical basis by Kolmogorov, has been applied to a variety of engineering studies. For instance, it has been used in the analysis of the size of randomly chosen gold particles from a collection of gold sand. A less obvious application of the law has been to a study of the stress release in earthquake fault zones (see Lomnitz, C., "Global Tectonics and Earthquake Risk," *Developments in Geotectonics*, Elsevier, Amsterdam, 1979).

Suppose that a sample of 10 grains of metallic sand taken from a large sand pile have respective lengths (in millimeters):

$$2.2,\ 3.4,\ 1.6,\ 0.8,\ 2.7,\ 3.3,\ 1.6,\ 2.8,\ 2.5,\ 1.9$$

Estimate the percentage of sand grains in the entire pile whose length is between 2 and 3 mm.

SOLUTION Taking the natural logarithm of these 10 data values, the following transformed data set results

$$.7885,\ 1.2238,\ .4700,\ -.2231,\ .9933,\ 1.1939,\ .4700,\ 1.0296,\ .9163,\ .6419$$

Because the sample mean and sample standard deviation of these data are

$$\overline{x} = .7504, \qquad s = .4351$$

it follows that the logarithm of the length of a randomly chosen grain has a normal distribution with mean approximately equal to .7504 and with standard deviation approximately equal to .4351. Hence, if X is the length of the grain, then

$$\begin{aligned} P\{2 < X < 3\} &= P\{\log(2) < \log(X) < \log(3)\} \\ &= P\left\{\frac{\log(2) - .7504}{.4351} < \frac{\log(X) - .7504}{.4351} < \frac{\log(3) - .7504}{.4351}\right\} \\ &= P\left\{-.1316 < \frac{\log(X) - .7504}{.4351} < .8003\right\} \\ &\approx \Phi(.8003) - \Phi(-.1316) \\ &= .3405 \quad \blacksquare \end{aligned}$$

In all of the foregoing examples, the maximum likelihood estimator of the population mean turned out to be the sample mean $\overline{X}$. To show that this is not always the situation, consider the following example.

EXAMPLE 7.2g (Estimating the Mean of a Uniform Distribution) Suppose $X_1, \ldots, X_n$ constitute a sample from a uniform distribution on $(0, \theta)$, where θ is unknown. Their joint density is thus

$$f(x_1, x_2, \ldots, x_n\, \theta) = \begin{cases} \dfrac{1}{\theta^n} & 0 < x_i < \theta, \quad i = 1, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

This density is maximized by choosing θ as small as possible. Since θ must be at least as large as all of the observed values x_i, it follows that the smallest possible choice of θ is equal to $\max(x_1, x_2, \ldots, x_n)$. Hence, the maximum likelihood estimator of θ is

$$\hat{\theta} = \max(X_1, X_2, \ldots, X_n)$$

It easily follows from the foregoing that the maximum likelihood estimator of $\theta/2$, the mean of the distribution, is $\max(X_1, X_2, \ldots, X_n)/2$. $\blacksquare$

Ignore This Section Below

*7.2.1 Estimating Life Distributions

Let X denote the age at death of a randomly chosen child born today. That is, $X = i$ if the newborn dies in its ith year, $i \geq 1$. To estimate the probability mass function of X, let λ_i denote the probability that a newborn who has survived his or her first $i - 1$ years

* Optional section.

Ignore the first part of this page, start with Section 7.3

To estimate s_i, the probability that a patient who has survived the first $i - 1$ months will also survive month i, we should take the fraction of those patients who began their ith month of drug taking and survived the month. For instance, because 11 of the 12 patients survived month 1, $\hat{s}_1 = 11/12$. Because all 11 patients who began month 2 survived, $\hat{s}_2 = 11/11$. Because 10 of the 11 patients who began month 3 survived, $\hat{s}_3 = 10/11$. Because 8 of the 9 patients who began their fourth month of taking the drug (all but the ones labelled 1, 3, and 3*) survived month 4, $\hat{s}_4 = 8/9$. Similar reasoning holds for the others, giving the following survival rate estimates:

$$\begin{aligned}
\hat{s}_1 &= 11/12 \\
\hat{s}_2 &= 11/11 \\
\hat{s}_3 &= 10/11 \\
\hat{s}_4 &= 8/9 \\
\hat{s}_5 &= 7/8 \\
\hat{s}_6 &= 7/7 \\
\hat{s}_7 &= 6/7 \\
\hat{s}_8 &= 4/5 \\
\hat{s}_9 &= 3/4 \\
\hat{s}_{10} &= 3/3 \\
\hat{s}_{11} &= 3/3 \\
\hat{s}_{12} &= 1/2 \\
\hat{s}_{13} &= 1/1 \\
\hat{s}_{14} &= 1/2
\end{aligned}$$

We can now use $\prod_{i=1}^{j} \hat{s}_i$ to estimate the probability that a drug taker survives at least j time periods, $j = 1, \ldots, 14$. For instance, our estimate of $P\{X > 6\}$ is 35/54.

7.3 INTERVAL ESTIMATES

Suppose that $X_1, \ldots, X_n$ is a sample from a normal population having unknown mean μ and known variance σ^2. It has been shown that $\overline{X} = \sum_{i=1}^{n} X_i/n$ is the maximum likelihood estimator for μ. However, we don't expect that the sample mean $\overline{X}$ will exactly equal μ, but rather that it will "be close." Hence, rather than a point estimate, it is sometimes more valuable to be able to specify an interval for which we have a certain degree of confidence that μ lies within. To obtain such an interval estimator, we make use of the probability distribution of the point estimator. Let us see how it works for the preceding situation.

In the foregoing, since the point estimator $\overline{X}$ is normal with mean μ and variance σ^2/n, it follows that

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} = \sqrt{n}\frac{(\overline{X}-\mu)}{\sigma}$$

has a standard normal distribution. Therefore,

$$P\left\{-1.96 < \sqrt{n}\frac{(\overline{X}-\mu)}{\sigma} < 1.96\right\} = .95$$

or, equivalently,

$$P\left\{-1.96\frac{\sigma}{\sqrt{n}} < \overline{X}-\mu < 1.96\frac{\sigma}{\sqrt{n}}\right\} = .95$$

Multiplying through by -1 yields the equivalent statement

$$P\left\{-1.96\frac{\sigma}{\sqrt{n}} < \mu-\overline{X} < 1.96\frac{\sigma}{\sqrt{n}}\right\} = .95$$

or, equivalently,

$$P\left\{\overline{X}-1.96\frac{\sigma}{\sqrt{n}} < \mu < \overline{X}+1.96\frac{\sigma}{\sqrt{n}}\right\} = .95$$

That is, 95 percent of the time μ will lie within $1.96\sigma/\sqrt{n}$ units of the sample average. If we now observe the sample and it turns out that $\overline{X} = \overline{x}$, then we say that "with 95 percent confidence"

$$\overline{x}-1.96\frac{\sigma}{\sqrt{n}} < \mu < \overline{x}+1.96\frac{\sigma}{\sqrt{n}} \tag{7.3.1}$$

That is, "with 95 percent confidence" we assert that the true mean lies within $1.96\sigma/\sqrt{n}$ of the observed sample mean. The interval

$$\left(\overline{x}-1.96\frac{\sigma}{\sqrt{n}}, \overline{x}+1.96\frac{\sigma}{\sqrt{n}}\right)$$

is called a *95 percent confidence interval estimate* of μ.

EXAMPLE 7.3a Suppose that when a signal having value μ is transmitted from location A the value received at location B is normally distributed with mean μ and variance 4. That is, if μ is sent, then the value received is $\mu + N$ where N, representing noise, is normal with mean 0 and variance 4. To reduce error, suppose the same value is sent 9 times. If the successive values received are 5, 8.5, 12, 15, 7, 9, 7.5, 6.5, 10.5, let us construct a 95 percent confidence interval for μ.

Since

$$\bar{x} = \frac{81}{9} = 9$$

It follows, under the assumption that the values received are independent, that a 95 percent confidence interval for μ is

$$\left(9 - 1.96\frac{\sigma}{3}, 9 + 1.96\frac{\sigma}{3}\right) = (7.69, 10.31)$$

Hence, we are "95 percent confident" that the true message value lies between 7.69 and 10.31. ■

The interval in Equation 7.3.1 is called a *two-sided confidence interval*. Sometimes, however, we are interested in determining a value so that we can assert with, say, 95 percent confidence, that μ is at least as large as that value.

To determine such a value, note that if Z is a standard normal random variable then

$$P\{Z < 1.645\} = .95$$

As a result,

$$P\left\{\sqrt{n}\frac{(\overline{X} - \mu)}{\sigma} < 1.645\right\} = .95$$

or

$$P\left\{\overline{X} - 1.645\frac{\sigma}{\sqrt{n}} < \mu\right\} = .95$$

Thus, a 95 percent *one-sided upper confidence interval* for μ is

$$\left(\bar{x} - 1.645\frac{\sigma}{\sqrt{n}}, \infty\right)$$

where $\bar{x}$ is the observed value of the sample mean.

A *one-sided lower confidence interval* is obtained similarly; when the observed value of the sample mean is $\bar{x}$, then the 95 percent one-sided lower confidence interval for μ is

$$\left(-\infty, \bar{x} + 1.645\frac{\sigma}{\sqrt{n}}\right)$$

EXAMPLE 7.3b Determine the upper and lower 95 percent confidence interval estimates of μ in Example 7.3a.

SOLUTION Since

$$1.645\frac{\sigma}{\sqrt{n}} = \frac{3.29}{3} = 1.097$$

the 95 percent upper confidence interval is

$$(9 - 1.097, \infty) = (7.903, \infty)$$

and the 95 percent lower confidence interval is

$$(-\infty, 9 + 1.097) = (-\infty, 10.097) \quad \blacksquare$$

We can also obtain confidence intervals of any specified level of confidence. To do so, recall that z_α is such that

$$P\{Z > z_\alpha\} = \alpha$$

when Z is a standard normal random variable. But this implies (see Figure 7.1) that for any α

$$P\{-z_{\alpha/2} < Z < z_{\alpha/2}\} = 1 - \alpha$$

As a result, we see that

$$P\left\{-z_{\alpha/2} < \sqrt{n}\frac{(\overline{X} - \mu)}{\sigma} < z_{\alpha/2}\right\} = 1 - \alpha$$

or

$$P\left\{-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \overline{X} - \mu < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\} = 1 - \alpha$$

or

$$P\left\{-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu - \overline{X} < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\} = 1 - \alpha$$

That is,

$$P\left\{\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\} = 1 - \alpha$$

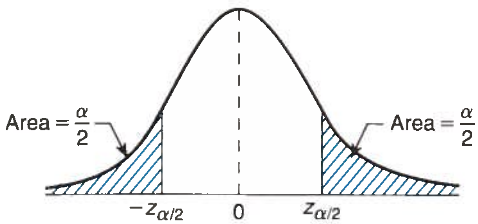


FIGURE 7.1 $P\{-z_{\alpha/2} < Z < z_{\alpha/2}\} = 1 - \alpha$.

Hence, a $100(1 - \alpha)$ percent two-sided confidence interval for μ is

$$\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \quad \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

where $\bar{x}$ is the observed sample mean.

Similarly, knowing that $Z = \sqrt{n}\frac{(\bar{X}-\mu)}{\sigma}$ is a standard normal random variable, along with the identities

$$P\{Z > z_\alpha\} = \alpha$$

and

$$P\{Z < -z_\alpha\} = \alpha$$

results in one-sided confidence intervals of any desired level of confidence. Specifically, we obtain that

$$\left(\bar{x} - z_\alpha\frac{\sigma}{\sqrt{n}}, \infty\right)$$

and

$$\left(-\infty, \bar{x} + z_\alpha\frac{\sigma}{\sqrt{n}}\right)$$

are, respectively, $100(1 - \alpha)$ percent one-sided upper and $100(1 - \alpha)$ percent one-sided lower confidence intervals for μ.

EXAMPLE 7.3c Use the data of Example 7.3a to obtain a 99 percent confidence interval estimate of μ, along with 99 percent one-sided upper and lower intervals.

SOLUTION Since $z_{.005} = 2.58$, and

$$2.58\frac{\alpha}{\sqrt{n}} = \frac{5.16}{3} = 1.72$$

it follows that a 99 percent confidence interval for μ is

$$9 \pm 1.72$$

That is, the 99 percent confidence interval estimate is (7.28, 10.72).

Also, since $z_{.01} = 2.33$, a 99 percent upper confidence interval is

$$(9 - 2.33(2/3), \infty) = (7.447, \infty)$$

Similarly, a 99 percent lower confidence interval is

$$(-\infty, 9 + 2.33(2/3)) = (-\infty, 10.553) \quad ■$$

Sometimes we are interested in a two-sided confidence interval of a certain level, say $1-\alpha$, and the problem is to choose the sample size n so that the interval is of a certain size. For instance, suppose that we want to compute an interval of length .1 that we can assert, with 99 percent confidence, contains μ. How large need n be? To solve this, note that as $z_{.005} = 2.58$ it follows that the 99 percent confidence interval for μ from a sample of size n is

$$\left(\overline{x} - 2.58\frac{\alpha}{\sqrt{n}}, \quad \overline{x} + 2.58\frac{\alpha}{\sqrt{n}}\right)$$

Hence, its length is

$$5.16\frac{\sigma}{\sqrt{n}}$$

Thus, to make the length of the interval equal to .1, we must choose

$$5.16\frac{\sigma}{\sqrt{n}} = .1$$

or

$$n = (51.6\sigma)^2$$

REMARK

The interpretation of "a $100(1-\alpha)$ percent confidence interval" can be confusing. It should be noted that we are *not* asserting that the probability that $\mu \in (\overline{x} - 1.96\sigma/\sqrt{n}, \overline{x} + 1.96\sigma/\sqrt{n})$ is .95, for there are no random variables involved in this assertion. What we are asserting is that the technique utilized to obtain this interval is such that 95 percent of the time that it is employed it will result in an interval in which μ lies. In other words, before the data are observed we can assert that with probability .95 the interval that will be obtained will contain μ, whereas after the data are obtained we can only assert that the resultant interval indeed contains μ "with confidence .95."

EXAMPLE 7.3d From past experience it is known that the weights of salmon grown at a commercial hatchery are normal with a mean that varies from season to season but with a standard deviation that remains fixed at 0.3 pounds. If we want to be 95 percent certain that our estimate of the present season's mean weight of a salmon is correct to within ± 0.1 pounds, how large a sample is needed?

SOLUTION A 95 percent confidence interval estimate for the unknown mean μ, based on a sample of size n, is

$$\mu \in \left(\overline{x} - 1.96\frac{\sigma}{\sqrt{n}}, \overline{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$$

Because the estimate $\bar{x}$ is within $1.96(\sigma/\sqrt{n}) = .588/\sqrt{n}$ of any point in the interval, it follows that we can be 95 percent certain that $\bar{x}$ is within 0.1 of μ provided that

$$\frac{.588}{\sqrt{n}} \leq 0.1$$

That is, provided that

$$\sqrt{n} \geq 5.88$$

or

$$n \geq 34.57$$

That is, a sample size of 35 or larger will suffice. ■

7.3.1 Confidence Interval for a Normal Mean When the Variance Is Unknown

Suppose now that $X_1, \ldots, X_n$ is a sample from a normal distribution with unknown mean μ and unknown variance σ^2, and that we wish to construct a $100(1-\alpha)$ percent confidence interval for μ. Since σ is unknown, we can no longer base our interval on the fact that $\sqrt{n}(\overline{X}-\mu)/\sigma$ is a standard normal random variable. However, by letting $S^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2/(n-1)$ denote the sample variance, then from Corollary 6.5.2 it follows that

$$\sqrt{n}\frac{(\overline{X} - \mu)}{S}$$

is a t-random variable with $n-1$ degrees of freedom. Hence, from the symmetry of the t-density function (see Figure 7.2), we have that for any $\alpha \in (0, 1/2)$,

$$P\left\{-t_{\alpha/2,n-1} < \sqrt{n}\frac{(\overline{X} - \mu)}{S} < t_{\alpha/2,n-1}\right\} = 1 - \alpha$$

or, equivalently,

$$P\left\{\overline{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right\} = 1 - \alpha$$

Thus, if it is observed that $\overline{X} = \bar{x}$ and $S = s$, then we can say that "with $100(1-\alpha)$ percent confidence"

$$\mu \in \left(\bar{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right)$$

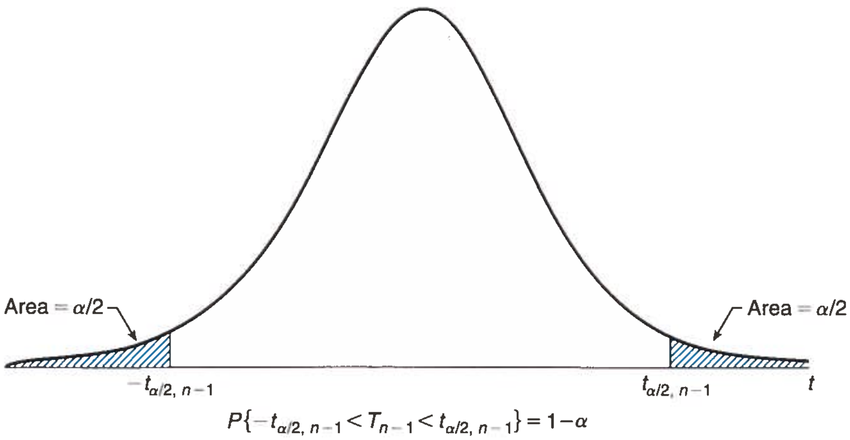


FIGURE 7.2 *t-density function.*

EXAMPLE 7.3e Let us again consider Example 7.3a but let us now suppose that when the value μ is transmitted at location A then the value received at location B is normal with mean μ and variance σ^2 but with σ^2 being unknown. If 9 successive values are, as in Example 7.3a, 5, 8.5, 12, 15, 7, 9, 7.5, 6.5, 10.5, compute a 95 percent confidence interval for μ.

SOLUTION A simple calculation yields that

$$\bar{x} = 9$$

and

$$s^2 = \frac{\sum x_i^2 - 9(\bar{x})^2}{8} = 9.5$$

or

$$s = 3.082$$

Hence, as $t_{.025,8} = 2.306$, a 95 percent confidence interval for μ is

$$\left[9 - 2.306\frac{(3.082)}{3}, 9 + 2.306\frac{(3.082)}{3}\right] = (6.63, 11.37)$$

a larger interval than obtained in Example 7.3a. The reason why the interval just obtained is larger than the one in Example 7.3a is twofold. The primary reason is that we have a larger estimated variance than in Example 7.3a. That is, in Example 7.3a we assumed that σ^2 was known to equal 4, whereas in this example we assumed it to be unknown and our estimate of it turned out to be 9.5, which resulted in a larger confidence interval. In fact, the confidence interval would have been larger than in Example 7.3a even if our estimate of σ^2 was again 4 because by having to estimate the variance we need to utilize the t-distribution, which has a greater variance and thus a larger spread than the standard normal (which can be used when σ^2 is assumed known). For instance, if it had turned out that $\bar{x} = 9$ and $s^2 = 4$, then our confidence interval would have been

$$(9 - 2.306 \cdot \tfrac{2}{3}, 9 + 2.306 \cdot \tfrac{2}{3}) = (7.46, 10.54)$$

which is larger than that obtained in Example 7.3a. ■

REMARKS

(a) The confidence interval for μ when σ is known is based on the fact that $\sqrt{n}(\overline{X} - \mu)/\sigma$ has a standard normal distribution. When σ is unknown, the foregoing approach is to estimate it by S and then use the fact that $\sqrt{n}(\overline{X} - \mu)/S$ has a t-distribution with $n - 1$ degrees of freedom.

(b) The length of a $100(1 - \alpha)$ percent confidence interval for μ is not always larger when the variance is unknown. For the length of such an interval is $2z_\alpha\sigma/\sqrt{n}$ when σ is known, whereas it is $2t_{\alpha,n-1}S/\sqrt{n}$ when σ is unknown; and it is certainly possible that the sample standard deviation S can turn out to be much smaller than σ. However, it can be shown that the mean length of the interval is longer when σ is unknown. That is, it can be shown that

$$t_{\alpha,n-1}E[S] \geq z_\alpha\sigma$$

Indeed, $E[S]$ is evaluated in Chapter 14 and it is shown, for instance, that

$$E[S] = \begin{cases} .94\sigma & \text{when } n = 5 \\ .97\sigma & \text{when } n = 9 \end{cases}$$

Since

$$z_{.025} = 1.96, \qquad t_{.025,4} = 2.78, \qquad t_{.025,8} = 2.31$$

the length of a 95 percent confidence interval from a sample of size 5 is $2 \times 1.96\sigma/\sqrt{5} = 1.75\sigma$ when σ is known, whereas its expected length is $2 \times 2.78 \times .94\sigma/\sqrt{5} = 2.34\sigma$ when σ is unknown — an increase of 33.7 percent. If the sample is of size 9, then the two values to compare are 1.31σ and 1.49σ — a gain of 13.7 percent. ■

A one-sided upper confidence interval can be obtained by noting that

$$P\left\{\sqrt{n}\frac{(\overline{X} - \mu)}{S} < t_{\alpha,n-1}\right\} = 1 - \alpha$$

or

$$P\left\{\overline{X} - \mu < \frac{S}{\sqrt{n}} t_{\alpha,n-1}\right\} = 1 - \alpha$$

or

$$P\left\{\mu > \overline{X} - \frac{S}{\sqrt{n}} t_{\alpha,n-1}\right\} = 1 - \alpha$$

Hence, if it is observed that $\overline{X} = \overline{x}$, $S = s$, then we can assert "with $100(1-\alpha)$ percent confidence" that

$$\mu \in \left(\overline{x} - \frac{s}{\sqrt{n}} t_{\alpha,n-1}, \infty\right)$$

Similarly, a $100(1-\alpha)$ lower confidence interval would be

$$\mu \in \left(-\infty, \overline{x} + \frac{s}{\sqrt{n}} t_{\alpha,n-1}\right)$$

Program 7.3.1 will compute both one- and two-sided confidence intervals for the mean of a normal distribution when the variance is unknown.

EXAMPLE 7.3f Determine a 95 percent confidence interval for the average resting pulse of the members of a health club if a random selection of 15 members of the club yielded the data 54, 63, 58, 72, 49, 92, 70, 73, 69, 104, 48, 66, 80, 64, 77. Also determine a 95 percent lower confidence interval for this mean.

SOLUTION We use Program 7.3.1 to obtain the solution (see Figure 7.3). ■

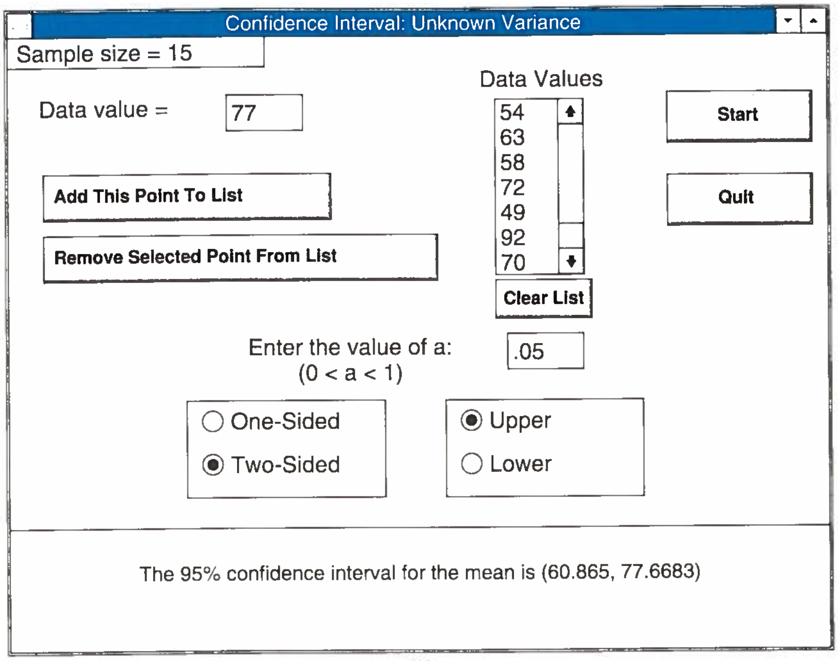
Our derivations of the $100(1-\alpha)$ percent confidence intervals for the population mean μ have assumed that the population distribution is normal. However, even when this is not the case, if the sample size is reasonably large then the intervals obtained will still be approximate $100(1-\alpha)$ percent confidence intervals for μ. This is true because, by the central limit theorem, $\sqrt{n}(\overline{X} - \mu)/\sigma$ will have approximately a normal distribution, and $\sqrt{n}(\overline{X} - \mu)/S$ will have approximately a t-distribution.

EXAMPLE 7.3g Simulation provides a powerful method for evaluating single and multi-dimensional integrals. For instance, let f be a function of an r-valued vector $(y_1, \ldots, y_r)$, and suppose that we want to estimate the quantity θ, defined by

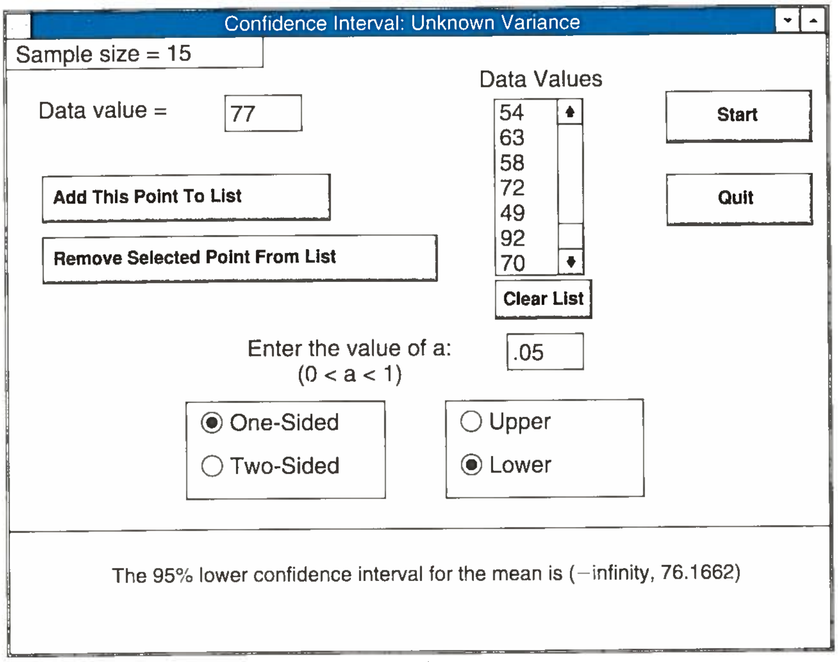
$$\theta = \int_0^1 \int_0^1 \cdots \int_0^1 f(y_1, y_2, \ldots, y_r)\, dy_1 dy_2, \ldots, dy_r$$

To accomplish this, note that if $U_1, U_2, \ldots, U_r$ are independent uniform random variables on (0, 1), then

$$\theta = E[f(U_1, U_2, \ldots, U_r)]$$



(a)



(b)

FIGURE 7.3 *(a) Two-sided and (b) lower 95 percent confidence intervals for Example 7.3f.*

Now, the values of independent uniform (0, 1) random variables can be approximated on a computer (by so-called *pseudo random numbers*); if we generate a vector of r of them, and evaluate f at this vector, then the value obtained, call it X_1, will be a random variable with mean θ. If we now repeat this process, then we obtain another value, call it X_2, which will have the same distribution as X_1. Continuing on, we can generate a sequence $X_1, X_2, \ldots, X_n$ of independent and identically distributed random variables with mean θ; we then use their observed values to estimate θ. This method of approximating integrals is called *Monte Carlo simulation*.

For instance, suppose we wanted to estimate the one-dimensional integral

$$\theta = \int_0^1 \sqrt{1-y^2}\, dy = E[\sqrt{1-U^2}]$$

where U is a uniform (0, 1) random variable. To do so, let $U_1, \ldots, U_{100}$ be independent uniform (0, 1) random variables, and set

$$X_i = \sqrt{1-U_i^2}, \qquad i = 1, \ldots, 100$$

In this way, we have generated a sample of 100 random variables having mean θ. Suppose that the computer generated values of $U_1, \ldots, U_{100}$, resulting in $X_1, \ldots, X_{100}$ having sample mean .786 and sample standard deviation .03. Consequently, since $t_{.025,99} = 1.985$, it follows that a 95 percent confidence interval for θ would be given by

$$.786 \pm 1.985(.003)$$

As a result, we could assert, with 95 percent confidence, that θ (which can be shown to equal $\pi/4$) is between .780 and .792. ■

7.3.2 Confidence Intervals for the Variance of a Normal Distribution

If $X_1, \ldots, X_n$ is a sample from a normal distribution having unknown parameters μ and σ^2, then we can construct a confidence interval for σ^2 by using the fact that

$$(n-1)\frac{S^2}{\sigma^2} \sim \chi^2_{n-1}$$

Hence,

$$P\left\{\chi^2_{1-\alpha/2,n-1} \leq (n-1)\frac{S^2}{\sigma^2} \leq \chi^2_{\alpha/2,n-1}\right\} = 1-\alpha$$

or, equivalently,

$$P\left\{\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}\right\} = 1-\alpha$$

Hence when $S^2 = s^2$, a $100(1-\alpha)$ percent confidence interval for σ^2 is

$$\left\{\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}\right\}$$

EXAMPLE 7.3h A standardized procedure is expected to produce washers with very small deviation in their thicknesses. Suppose that 10 such washers were chosen and measured. If the thicknesses of these washers were, in inches,

.123	.133
.124	.125
.126	.128
.120	.124
.130	.126

what is a 90 percent confidence interval for the standard deviation of the thickness of a washer produced by this procedure?

SOLUTION A computation gives that

$$S^2 = 1.366 \times 10^{-5}$$

Because $\chi^2_{.05,9} = 16.917$ and $\chi^2_{.95,9} = 3.334$, and because

$$\frac{9 \times 1.366 \times 10^{-5}}{16.917} = 7.267 \times 10^{-6}, \qquad \frac{9 \times 1.366 \times 10^{-5}}{3.334} = 36.875 \times 10^{-6}$$

TABLE 7.1 *100(1 − α) Percent Confidence Intervals*

$$X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$$

$$\overline{X} = \sum_{i=1}^{n} X_i/n, \qquad S = \sqrt{\sum_{i=1}^{n} (X_i - \overline{X})^2/(n-1)}$$

Assumption	Parameter	Confidence Interval	Lower Interval	Upper Interval
σ^2 known	μ	$\overline{X} \pm z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$	$\left(-\infty, \overline{X} + z_\alpha \dfrac{\sigma}{\sqrt{n}}\right)$	$\left(\overline{X} + z_\alpha \dfrac{\sigma}{\sqrt{n}}, \infty\right)$
σ^2 unknown	μ	$\overline{X} \pm t_{\alpha/2,n-1}\dfrac{S}{\sqrt{n}}$	$\left(-\infty, \overline{X} + t_{\alpha,n-1} \dfrac{S}{\sqrt{n}}\right)$	$\left(\overline{X} - t_{\alpha,n-1} \dfrac{S}{\sqrt{n}}, \infty\right)$
μ unknown	σ^2	$\left(\dfrac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}}, \dfrac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}\right)$	$\left(0, \dfrac{(n-1)S^2}{\chi^2_{1-\alpha,n-1}}\right)$	$\left(\dfrac{(n-1)S^2}{\chi^2_{\alpha,n-1}}, \infty\right)$

it follows that, with confidence .90,

$$\sigma^2 \in (7.267 \times 10^{-6}, \quad 36.875 \times 10^{-6})$$

Taking square roots yields that, with confidence .90,

$$\sigma \in (2.696 \times 10^{-3}, \quad 6.072 \times 10^{-3}) \quad \blacksquare$$

One-sided confidence intervals for σ^2 are obtained by similar reasoning and are presented in Table 7.1, which sums up the results of this section.

Ignore this section below

7.4 ESTIMATING THE DIFFERENCE IN MEANS OF TWO NORMAL POPULATIONS

Let $X_1, X_2, \ldots, X_n$ be a sample of size n from a normal population having mean μ_1 and variance σ_1^2 and let $Y_1, \ldots, Y_m$ be a sample of size m from a different normal population having mean μ_2 and variance σ_2^2 and suppose that the two samples are independent of each other. We are interested in estimating $\mu_1 - \mu_2$.

Since $\overline{X} = \sum_{i=1}^{n} X_i/n$ and $\overline{Y} = \sum_{i=1}^{m} Y_i/m$ are the maximum likelihood estimators of μ_1 and μ_2 it seems intuitive (and can be proven) that $\overline{X} - \overline{Y}$ is the maximum likelihood estimator of $\mu_1 - \mu_2$.

To obtain a confidence interval estimator, we need the distribution of $\overline{X} - \overline{Y}$. Because

$$\overline{X} \sim \mathcal{N}(\mu_1, \sigma_1^2/n)$$
$$\overline{Y} \sim \mathcal{N}(\mu_2, \sigma_2^2/m)$$

it follows from the fact that the sum of independent normal random variables is also normal, that

$$\overline{X} - \overline{Y} \sim \mathcal{N}\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right)$$

Hence, assuming σ_1^2 and σ_2^2 are known, we have that

$$\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}}} \sim \mathcal{N}(0, 1) \tag{7.4.1}$$

and so

$$P\left\{-z_{\alpha/2} < \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}}} < z_{\alpha/2}\right\} = 1 - \alpha$$

Ignore this first part below, start with Section 7.8

To determine the constant c resulting in minimal mean square error, we differentiate to obtain

$$\frac{d}{dc}r(d_c(\mathbf{X}),\theta) = \frac{2cn\theta^2}{(n+2)(n+1)^2} + \frac{2\theta^2 n}{n+1}\left(\frac{cn}{n+1} - 1\right)$$

Equating this to 0 shows that the best constant c — call it c^* — is such that

$$\frac{c^*}{n+2} + c^*n - (n+1) = 0$$

or

$$c^* = \frac{(n+1)(n+2)}{n^2+2n+1} = \frac{n+2}{n+1}$$

Substituting this value of c into Equation 7.7.4 yields that

$$\begin{aligned} r\left(\frac{n+2}{n+1}\max_i X_i, \theta\right) &= \frac{(n+2)n\theta^2}{(n+1)^4} + \theta^2\left(\frac{n(n+2)}{(n+1)^2} - 1\right)^2 \\ &= \frac{(n+2)n\theta^2}{(n+1)^4} + \frac{\theta^2}{(n+1)^4} \\ &= \frac{\theta^2}{(n+1)^2} \end{aligned}$$

A comparison with Equation 7.7.3 shows that the (biased) estimator $(n+2)/(n+1)\max_i X_i$ has about half the mean square error of the maximum likelihood estimator $\max_i X_i$. ■

*7.8 THE BAYES ESTIMATOR

In certain situations it seems reasonable to regard an unknown parameter θ as being the value of a random variable from a given probability distribution. This usually arises when, prior to the observance of the outcomes of the data $X_1, \ldots, X_n$, we have some information about the value of θ and this information is expressible in terms of a probability distribution (called appropriately the *prior* distribution of θ). For instance, suppose that from past experience we know that θ is equally likely to be near any value in the interval (0, 1). Hence, we could reasonably assume that θ is chosen from a uniform distribution on (0, 1).

Suppose now that our prior feelings about θ are that it can be regarded as being the value of a continuous random variable having probability density function $p(\theta)$; and suppose that we are about to observe the value of a sample whose distribution depends on θ. Specifically, suppose that $f(x|\theta)$ represents the likelihood — that is, it is the probability mass function in the discrete case or the probability density function in the continuous

* Optional section.

case — that a data value is equal to x when θ is the value of the parameter. If the observed data values are $X_i = x_i, i = 1, \ldots, n$, then the updated, or conditional, probability density function of θ is as follows:

$$\begin{aligned} f(\theta|x_1, \ldots, x_n) &= \frac{f(\theta, x_1, \ldots, x_n)}{f(x_1, \ldots, x_n)} \\ &= \frac{p(\theta)f(x_1, \ldots, x_n|\theta)}{\int f(x_1, \ldots, x_n|\theta)p(\theta)\, d\theta} \end{aligned}$$

The conditional density function $f(\theta|x_1, \ldots, x_n)$ is called the *posterior* density function. (Thus, before observing the data, one's feelings about θ are expressed in terms of the prior distribution, whereas once the data are observed, this prior distribution is updated to yield the posterior distribution.)

Now we have shown that whenever we are given the probability distribution of a random variable, the best estimate of the value of that random variable, in the sense of minimizing the expected squared error, is its mean. Therefore, it follows that the best estimate of θ, given the data values $X_i = x_i, i = 1, \ldots, n$, is the mean of the posterior distribution $f(\theta|x_1, \ldots, x_n)$. This estimator, called the *Bayes estimator*, is written as $E[\theta|X_1, \ldots, X_n]$. That is, if $X_i = x_i$, $i = 1, \ldots, n$, then the value of the Bayes estimator is

$$E[\theta|X_1 = x_1, \ldots, X_n = x_n] = \int \theta f(\theta|x_1, \ldots, x_n)\, d\theta$$

EXAMPLE 7.8a Suppose that $X_1, \ldots, X_n$ are independent Bernoulli random variables, each having probability mass function given by

$$f(x|\theta) = \theta^x(1-\theta)^{1-x}, \qquad x = 0, 1$$

where θ is unknown. Further, suppose that θ is chosen from a uniform distribution on (0, 1). Compute the Bayes estimator of θ.

SOLUTION We must compute $E[\theta|X_1, \ldots, X_n]$. Since the prior density of θ is the uniform density

$$p(\theta) = 1, \qquad 0 < \theta < 1$$

we have that the conditional density of θ given $X_1, \ldots, X_n$ is given by

$$\begin{aligned} f(\theta|x_1, \ldots, x_n) &= \frac{f(x_1, \ldots, x_n, \theta)}{f(x_1, \ldots, x_n)} \\ &= \frac{f(x_1, \ldots, x_n|\theta)p(\theta)}{\int_0^1 f(x_1, \ldots, x_n|\theta)p(\theta)\, d\theta} \\ &= \frac{\theta^{\Sigma_1^n x_i}(1-\theta)^{n-\Sigma_1^n x_i}}{\int_0^1 \theta^{\Sigma_1^n x_i}(1-\theta)^{n-\Sigma_1^n x_i}\, d\theta} \end{aligned}$$

Now it can be shown that for integral values m and r

$$\int_0^1 \theta^m (1-\theta)^r \, d\theta = \frac{m!r!}{(m+r+1)!} \tag{7.8.1}$$

Hence, upon letting $x = \sum_{i=1}^n x_i$

$$f(\theta|x_1, \ldots, x_n) = \frac{(n+1)!\theta^x (1-\theta)^{n-x}}{x!(n-x)!} \tag{7.8.2}$$

Therefore,

$$\begin{aligned} E[\theta|x_1, \ldots, x_n] &= \frac{(n+1)!}{x!(n-x)!} \int_0^1 \theta^{1+x}(1-\theta)^{n-x} \, d\theta \\ &= \frac{(n+1)!}{x!(n-x)!} \frac{(1+x)!(n-x)!}{(n+2)!} \qquad \text{from Equation 7.8.1} \\ &= \frac{x+1}{n+2} \end{aligned}$$

Thus, the Bayes estimator is given by

$$E[\theta|X_1, \ldots, X_n] = \frac{\sum_{i=1}^n X_i + 1}{n+2}$$

As an illustration, if 10 independent trials, each of which results in a success with probability θ, result in 6 successes, then assuming a uniform (0, 1) prior distribution on θ, the Bayes estimator of θ is 7/12 (as opposed, for instance, to the maximum likelihood estimator of 6/10). ■

REMARK

The conditional distribution of θ given that $X_i = x_i, i = 1, \ldots, n$, whose density function is given by Equation 7.8.2, is called the beta distribution with parameters $\sum_{i=1}^n x_i + 1$, $n - \sum_{i=1}^n x_i + 1$. ■

EXAMPLE 7.8b Suppose $X_1, \ldots, X_n$ are independent normal random variables, each having unknown mean θ and known variance σ_0^2. If θ is itself selected from a normal population having known mean μ and known variance σ^2, what is the Bayes estimator of θ?

SOLUTION In order to determine $E[\theta|X_1, \ldots, X_n]$, the Bayes estimator, we need first determine the conditional density of θ given the values of $X_1, \ldots, X_n$. Now

$$f(\theta|x_1, \ldots, x_n) = \frac{f(x_1, \ldots, x_n|\theta)p(\theta)}{f(x_1, \ldots, x_n)}$$

where

$$f(x_1, \ldots, x_n|\theta) = \frac{1}{(2\pi)^{n/2}\sigma_0^n} \exp\left\{-\sum_{i=1}^{n}(x_i - \theta)^2/2\sigma_0^2\right\}$$

$$p(\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-(\theta - \mu)^2/2\sigma^2\}$$

and

$$f(x_1, \ldots, x_n) = \int_{-\infty}^{\infty} f(x_1, \ldots, x_n|\theta)p(\theta)\, d\theta$$

With the help of a little algebra, it can now be shown that this conditional density is a *normal* density with mean

$$\begin{aligned} E[\theta|X_1, \ldots, X_n] &= \frac{n\sigma^2}{n\sigma^2 + \sigma_0^2}\overline{X} + \frac{\sigma_0^2}{n\sigma^2 + \sigma_0^2}\mu \qquad (7.8.3) \\ &= \frac{\dfrac{n}{\sigma_0^2}}{\dfrac{n}{\sigma_0^2} + \dfrac{1}{\sigma^2}}\overline{X} + \frac{\dfrac{1}{\sigma^2}}{\dfrac{n}{\sigma_0^2} + \dfrac{1}{\sigma^2}}\mu \end{aligned}$$

and variance

$$\mathrm{Var}(\theta|X_1, \ldots, X_n) = \frac{\sigma_0^2\sigma^2}{n\sigma^2 + \sigma_0^2}$$

Writing the Bayes estimator as we did in Equation 7.8.3 is informative, for it shows that it is a weighted average of $\overline{X}$, the sample mean, and μ, the *a priori* mean. In fact, the weights given to these two quantities are in proportion to the inverses of σ_0^2/n (the conditional variance of the sample mean $\overline{X}$ given θ) and σ^2 (the variance of the prior distribution). ■

REMARK: ON CHOOSING A NORMAL PRIOR

As illustrated by Example 7.8b, it is computationally very convenient to choose a normal prior for the unknown mean θ of a normal distribution — for then the Bayes estimator is simply given by Equation 7.8.3. This raises the question of how one should go about determining whether there is a normal prior that reasonably represents one's prior feelings about the unknown mean.

To begin, it seems reasonable to determine the value — call it μ — that you *a priori* feel is most likely to be near θ. That is, we start with the mode (which equals the mean when the distribution is normal) of the prior distribution. We should then try to ascertain whether or not we believe that the prior distribution is symmetric about μ. That is, for each $a > 0$ do we believe that it is just as likely that θ will lie between $\mu - a$ and μ as it is that it will be between μ and $\mu + a$? If the answer is positive, then we accept, as a working hypothesis, that our prior feelings about θ can be expressed in terms of a prior distribution that is normal with mean μ. To determine σ, the standard deviation of the normal prior, think of an interval centered about μ that you *a priori* feel is 90 percent certain to contain θ. For instance, suppose you feel 90 percent (no more and no less) certain that θ will lie between $\mu - a$ and $\mu + a$. Then, since a normal random variable θ with mean μ and variance σ^2 is such that

$$P\left\{-1.645 < \frac{\theta - \mu}{\sigma} < 1.645\right\} = .90$$

or

$$P\{\mu - 1.645\sigma < \theta < \mu + 1.645\sigma\} = .90$$

it seems reasonable to take

$$1.645\sigma = a \quad \text{or} \quad \sigma = \frac{a}{1.645}$$

Thus, if your prior feelings can indeed be reasonably described by a normal distribution, then that distribution would have mean μ and standard deviation $\sigma = a/1.645$. As a test of whether this distribution indeed fits your prior feelings you might ask yourself such questions as whether you are 95 percent certain that θ will fall between $\mu - 1.96\sigma$ and $\mu + 1.96\sigma$, or whether you are 99 percent certain that θ will fall between $\mu - 2.58\sigma$ and $\mu + 2.58\sigma$, where these intervals are determined by the equalities

$$P\left\{-1.96 < \frac{\theta - \mu}{\sigma} < 1.96\right\} = .95$$
$$P\left\{-2.58 < \frac{\theta - \mu}{\sigma} < 2.58\right\} = .99$$

which hold when θ is normal with mean μ and variance σ^2.

EXAMPLE 7.8c Consider the likelihood function $f(x_1, \ldots, x_n|\theta)$ and suppose that θ is uniformly distributed over some interval (a, b). The posterior density of θ given $X_1, \ldots, X_n$ equals

$$\begin{aligned} f(\theta|x_1, \ldots, x_n) &= \frac{f(x_1, \ldots, x_n|\theta)p(\theta)}{\int_a^b f(x_1, \ldots, x_n|\theta)p(\theta)\, d\theta} \\ &= \frac{f(x_1, \ldots, x_n|\theta)}{\int_a^b f(x_1, \ldots, x_n|\theta)\, d\theta} \qquad a < \theta < b \end{aligned}$$

Now the *mode* of a density $f(\theta)$ was defined to be that value of θ that maximizes $f(\theta)$. By the foregoing, it follows that the mode of the density $f(\theta|x_1, \ldots, x_n)$ is that value of θ

maximizing $f(x_1, \ldots, x_n|\theta)$; that is, it is just the maximum likelihood estimate of θ [when it is constrained to be in (a, b)]. In other words, the maximum likelihood estimate equals the mode of the posterior distribution when a uniform prior distribution is assumed. ■

If, rather than a point estimate, we desire an interval in which θ lies with a specified probability — say $1 - \alpha$ — we can accomplish this by choosing values a and b such that

$$\int_a^b f(\theta|x_1, \ldots, x_n)\, d\theta = 1 - \alpha$$

EXAMPLE 7.8d Suppose that if a signal of value s is sent from location A, then the signal value received at location B is normally distributed with mean s and variance 60. Suppose also that the value of a signal sent at location A is, *a priori*, known to be normally distributed with mean 50 and variance 100. If the value received at location B is equal to 40, determine an interval that will contain the actual value sent with probability .90.

SOLUTION It follows from Example 7.8b that the conditional distribution of S, the signal value sent, given that 40 is the value received, is normal with mean and variance given by

$$E[S|\text{data}] = \frac{1/60}{1/60 + 1/100}40 + \frac{1/100}{1/60 + 1/100}50 = 43.75$$

$$\text{Var}(S|\text{data}) = \frac{1}{1/60 + 1/100} = 37.5$$

Hence, given that the value received is 40, $(S - 43.75)/\sqrt{37.5}$ has a unit standard distribution and so

$$P\left\{-1.645 < \frac{S - 43.75}{\sqrt{37.5}} < 1.645|\text{data}\right\} = .90$$

or

$$P\{43.75 - 1.645\sqrt{37.5} < S < 43.75 + 1.645\sqrt{37.5}|\text{data}\} = .95$$

That is, with *probability* .90, the true signal sent lies within the interval (33.68, 53.82). ■

Don't do the problems, unless you really want to

Problems

1. Let $X_1, \ldots, X_n$ be a sample from the distribution whose density function is

$$f(x) = \begin{cases} e^{-(x-\theta)} & x \geq \theta \\ 0 & \text{otherwise} \end{cases}$$

Determine the maximum likelihood estimator of θ.

Chapter 8

HYPOTHESIS TESTING

8.1 INTRODUCTION

As in the previous chapter, let us suppose that a random sample from a population distribution, specified except for a vector of unknown parameters, is to be observed. However, rather than wishing to explicitly estimate the unknown parameters, let us now suppose that we are primarily concerned with using the resulting sample to test some particular hypothesis concerning them. As an illustration, suppose that a construction firm has just purchased a large supply of cables that have been guaranteed to have an average breaking strength of at least 7,000 psi. To verify this claim, the firm has decided to take a random sample of 10 of these cables to determine their breaking strengths. They will then use the result of this experiment to ascertain whether or not they accept the cable manufacturer's hypothesis that the population mean is at least 7,000 pounds per square inch.

A statistical hypothesis is usually a statement about a set of parameters of a population distribution. It is called a hypothesis because it is not known whether or not it is true. A primary problem is to develop a procedure for determining whether or not the values of a random sample from this population are consistent with the hypothesis. For instance, consider a particular normally distributed population having an unknown mean value θ and known variance 1. The statement "θ is less than 1" is a statistical hypothesis that we could try to test by observing a random sample from this population. If the random sample is deemed to be consistent with the hypothesis under consideration, we say that the hypothesis has been "accepted"; otherwise we say that it has been "rejected."

Note that in accepting a given hypothesis we are not actually claiming that it is true but rather we are saying that the resulting data appear to be consistent with it. For instance, in the case of a normal $(\theta, 1)$ population, if a resulting sample of size 10 has an average value of 1.25, then although such a result cannot be regarded as being evidence in favor of the hypothesis "$\theta < 1$," it is not inconsistent with this hypothesis, which would thus be accepted. On the other hand, if the sample of size 10 has an average value of 3, then even though a sample value that large is possible when $\theta < 1$, it is so unlikely that it seems inconsistent with this hypothesis, which would thus be rejected.

8.2 SIGNIFICANCE LEVELS

Consider a population having distribution F_θ, where θ is unknown, and suppose we want to test a specific hypothesis about θ. We shall denote this hypothesis by H_0 and call it the *null hypothesis*. For example, if F_θ is a normal distribution function with mean θ and variance equal to 1, then two possible null hypotheses about θ are

$$\textbf{(a)}\ H_0 : \theta = 1$$
$$\textbf{(b)}\ H_0 : \theta \leq 1$$

Thus the first of these hypotheses states that the population is normal with mean 1 and variance 1, whereas the second states that it is normal with variance 1 and a mean less than or equal to 1. Note that the null hypothesis in (a), when true, completely specifies the population distribution, whereas the null hypothesis in (b) does not. A hypothesis that, when true, completely specifies the population distribution is called a *simple* hypothesis; one that does not is called a *composite* hypothesis.

Suppose now that in order to test a specific null hypothesis H_0, a population sample of size n — say $X_1, \ldots, X_n$ — is to be observed. Based on these n values, we must decide whether or not to accept H_0. A test for H_0 can be specified by defining a region C in n-dimensional space with the proviso that the hypothesis is to be rejected if the random sample $X_1, \ldots, X_n$ turns out to lie in C and accepted otherwise. The region C is called the *critical region*. In other words, the statistical test determined by the critical region C is the one that

$$\text{accepts} \quad H_0 \quad \text{if} \quad (X_1, X_2, \ldots, X_n) \notin C$$

and

$$\text{rejects} \quad H_0 \quad \text{if} \quad (X_1, \ldots, X_n) \in C$$

For instance, a common test of the hypothesis that θ, the mean of a normal population with variance 1, is equal to 1 has a critical region given by

$$C = \left\{ (X_1, \ldots, X_n) : \left| \frac{\sum_{i=1}^{n} X_i}{n} - 1 \right| > \frac{1.96}{\sqrt{n}} \right\} \tag{8.2.1}$$

Thus, this test calls for rejection of the null hypothesis that $\theta = 1$ when the sample average differs from 1 by more than 1.96 divided by the square root of the sample size.

It is important to note when developing a procedure for testing a given null hypothesis H_0 that, in any test, two different types of errors can result. The first of these, called a *type I error*, is said to result if the test incorrectly calls for rejecting H_0 when it is indeed correct. The second, called a *type II error*, results if the test calls for accepting H_0 when it is false.

Now, as was previously mentioned, the objective of a statistical test of H_0 is not to explicitly determine whether or not H_0 is true but rather to determine if its validity is consistent with the resultant data. Hence, with this objective it seems reasonable that H_0 should only be rejected if the resultant data are very unlikely when H_0 is true. The classical way of accomplishing this is to specify a value α and then require the test to have the property that whenever H_0 is true its probability of being rejected is never greater than α. The value α, called the *level of significance of the test*, is usually set in advance, with commonly chosen values being $\alpha = .1, .05, .005$. In other words, the classical approach to testing H_0 is to fix a significance level α and then require that the test have the property that the probability of a type I error occurring can never be greater than α.

Suppose now that we are interested in testing a certain hypothesis concerning θ, an unknown parameter of the population. Specifically, for a given set of parameter values w, suppose we are interested in testing

$$H_0 : \theta \in w$$

A common approach to developing a test of H_0, say at level of significance α, is to start by determining a point estimator of θ — say $d(\mathbf{X})$. The hypothesis is then rejected if $d(\mathbf{X})$ is "far away" from the region w. However, to determine how "far away" it need be to justify rejection of H_0, we need to determine the probability distribution of $d(\mathbf{X})$ when H_0 is true since this will usually enable us to determine the appropriate critical region so as to make the test have the required significance level α. For example, the test of the hypothesis that the mean of a normal $(\theta, 1)$ population is equal to 1, given by Equation 8.2.1, calls for rejection when the point estimate of θ — that is, the sample average — is farther than $1.96/\sqrt{n}$ away from 1. As we will see in the next section, the value $1.96/\sqrt{n}$ was chosen to meet a level of significance of $\alpha = .05$.

8.3 TESTS CONCERNING THE MEAN OF A NORMAL POPULATION

8.3.1 Case of Known Variance

Suppose that $X_1, \ldots, X_n$ is a sample of size n from a normal distribution having an unknown mean μ and a known variance σ^2 and suppose we are interested in testing the null hypothesis

$$H_0 : \mu = \mu_0$$

against the alternative hypothesis

$$H_1 : \mu \neq \mu_0$$

where μ_0 is some specified constant.

Since $\overline{X} = \sum_{i=1}^{n} X_i/n$ is a natural point estimator of μ, it seems reasonable to accept H_0 if $\overline{X}$ is not too far from μ_0. That is, the critical region of the test would be of the form

$$C = \{X_1, \ldots, X_n : |\overline{X} - \mu_0| > c\} \tag{8.3.1}$$

for some suitably chosen value c.

If we desire that the test has significance level α, then we must determine the critical value c in Equation 8.3.1 that will make the type I error equal to α. That is, c must be such that

$$P_{\mu_0}\{|\overline{X} - \mu_0| > c\} = \alpha \tag{8.3.2}$$

where we write P_{μ_0} to mean that the preceding probability is to be computed under the assumption that $\mu = \mu_0$. However, when $\mu = \mu_0, \overline{X}$ will be normally distributed with mean μ_0 and variance σ^2/n and so Z, defined by

$$Z \equiv \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$$

will have a standard normal distribution. Now Equation 8.3.2 is equivalent to

$$P\left\{|Z| > \frac{c\sqrt{n}}{\sigma}\right\} = \alpha$$

or, equivalently,

$$2P\left\{Z > \frac{c\sqrt{n}}{\sigma}\right\} = \alpha$$

where Z is a standard normal random variable. However, we know that

$$P\{Z > z_{\alpha/2}\} = \alpha/2$$

and so

$$\frac{c\sqrt{n}}{\sigma} = z_{\alpha/2}$$

or

$$c = \frac{z_{\alpha/2}\sigma}{\sqrt{n}}$$

Thus, the significance level α test is to reject H_0 if $|\overline{X} - \mu_0| > z_{\alpha/2}\sigma/\sqrt{n}$ and accept otherwise; or, equivalently, to

$$\begin{aligned} &\text{reject} \quad H_0 \quad \text{if} \quad \frac{\sqrt{n}}{\sigma}|\overline{X} - \mu_0| > z_{\alpha/2} \\ &\text{accept} \quad H_0 \quad \text{if} \quad \frac{\sqrt{n}}{\sigma}|\overline{X} - \mu_0| \leq z_{\alpha/2} \end{aligned} \tag{8.3.3}$$

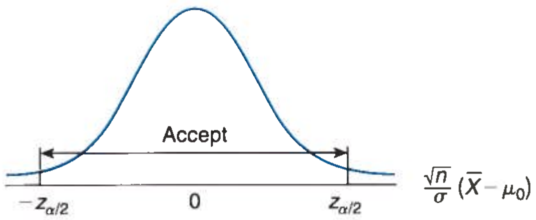


FIGURE 8.1

This can be pictorially represented as shown in Figure 8.1, where we have superimposed the standard normal density function [which is the density of the test statistic $\sqrt{n}(\overline{X} - \mu_0)/\sigma$ when H_0 is true].

EXAMPLE 8.3a It is known that if a signal of value μ is sent from location A, then the value received at location B is normally distributed with mean μ and standard deviation 2. That is, the random noise added to the signal is an $N(0, 4)$ random variable. There is reason for the people at location B to suspect that the signal value $\mu = 8$ will be sent today. Test this hypothesis if the same signal value is independently sent five times and the average value received at location B is $\overline{X} = 9.5$.

SOLUTION Suppose we are testing at the 5 percent level of significance. To begin, we compute the test statistic

$$\frac{\sqrt{n}}{\sigma}|\overline{X} - \mu_0| = \frac{\sqrt{5}}{2}(1.5) = 1.68$$

Since this value is less than $z_{.025} = 1.96$, the hypothesis is accepted. In other words, the data are not inconsistent with the null hypothesis in the sense that a sample average as far from the value 8 as observed would be expected, when the true mean is 8, over 5 percent of the time. Note, however, that if a less stringent significance level were chosen — say $\alpha = .1$ — then the null hypothesis would have been rejected. This follows since $z_{.05} = 1.645$, which is less than 1.68. Hence, if we would have chosen a test that had a 10 percent chance of rejecting H_0 when H_0 was true, then the null hypothesis would have been rejected.

The "correct" level of significance to use in a given situation depends on the individual circumstances involved in that situation. For instance, if rejecting a null hypothesis H_0 would result in large costs that would thus be lost if H_0 were indeed true, then we might elect to be quite conservative and so choose a significance level of .05 or .01. Also, if we initially feel strongly that H_0 was correct, then we would require very stringent data evidence to the contrary for us to reject H_0. (That is, we would set a very low significance level in this situation.) ■

The test given by Equation 8.3.3 can be described as follows: For any observed value of the test statistic $\sqrt{n}|\overline{X} - \mu_0|/\sigma$, call it v, the test calls for rejection of the null hypothesis if the probability that the test statistic would be as large as v when H_0 is true is less than or equal to the significance level α. From this, it follows that we can determine whether or not to accept the null hypothesis by computing, first, the value of the test statistic and, second, the probability that a unit normal would (in absolute value) exceed that quantity. This probability — called the p-value of the test — gives the critical significance level in the sense that H_0 will be accepted if the significance level α is less than the p-value and rejected if it is greater than or equal.

In practice, the significance level is often not set in advance but rather the data are looked at to determine the resultant p-value. Sometimes, this critical significance level is clearly much larger than any we would want to use, and so the null hypothesis can be readily accepted. At other times the p-value is so small that it is clear that the hypothesis should be rejected.

EXAMPLE 8.3b In Example 8.3a, suppose that the average of the 5 values received is $\overline{X} = 8.5$. In this case,

$$\frac{\sqrt{n}}{\sigma}|\overline{X} - \mu_0| = \frac{\sqrt{5}}{4} = .559$$

Since

$$\begin{aligned} P\{|Z| > .559\} &= 2P\{Z > .559\} \\ &= 2 \times .288 = .576 \end{aligned}$$

it follows that the p-value is .576 and thus the null hypothesis H_0 that the signal sent has value 8 would be accepted at any significance level $\alpha < .576$. Since we would clearly never want to test a null hypothesis using a significance level as large as .576, H_0 would be accepted.

On the other hand, if the average of the data values were 11.5, then the p-value of the test that the mean is equal to 8 would be

$$\begin{aligned} P\{|Z| > 1.75\sqrt{5}\} &= P\{|Z| > 3.913\} \\ &\approx .00005 \end{aligned}$$

For such a small p-value, the hypothesis that the value 8 was sent is rejected. ■

We have not yet talked about the probability of a type II error — that is, the probability of accepting the null hypothesis when the true mean μ is unequal to μ_0. This probability will depend on the value of μ, and so let us define $\beta(\mu)$ by

$$\begin{aligned}
\beta(\mu) &= P_\mu\{\text{acceptance of } H_0\} \\
&= P_\mu\left\{\left|\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\right| \le z_{\alpha/2}\right\} \\
&= P_\mu\left\{-z_{\alpha/2} \le \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \le z_{\alpha/2}\right\}
\end{aligned}$$

The function $\beta(\mu)$ is called the *operating characteristic* (or OC) *curve* and represents the probability that H_0 will be accepted when the true mean is μ.

To compute this probability, we use the fact that $\overline{X}$ is normal with mean μ and variance σ^2/n and so

$$Z \equiv \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$$

Hence,

$$\begin{aligned}
\beta(\mu) &= P_\mu\left\{-z_{\alpha/2} \le \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \le z_{\alpha/2}\right\} \\
&= P_\mu\left\{-z_{\alpha/2} - \frac{\mu}{\sigma/\sqrt{n}} \le \frac{\overline{X}-\mu_0-\mu}{\sigma/\sqrt{n}} \le z_{\alpha/2} - \frac{\mu}{\sigma/\sqrt{n}}\right\} \\
&= P_\mu\left\{-z_{\alpha/2} - \frac{\mu}{\sigma/\sqrt{n}} \le Z - \frac{\mu_0}{\sigma/\sqrt{n}} \le z_{\alpha/2} - \frac{\mu}{\sigma/\sqrt{n}}\right\} \\
&= P\left\{\frac{\mu_0-\mu}{\sigma/\sqrt{n}} - z_{\alpha/2} \le Z \le \frac{\mu_0-\mu}{\sigma/\sqrt{n}} + z_{\alpha/2}\right\} \\
&= \Phi\left(\frac{\mu_0-\mu}{\sigma/\sqrt{n}} + z_{\alpha/2}\right) - \Phi\left(\frac{\mu_0-\mu}{\sigma/\sqrt{n}} - z_{\alpha/2}\right) \qquad (8.3.4)
\end{aligned}$$

where Φ is the standard normal distribution function.

For a fixed significance level α, the OC curve given by Equation 8.3.4 is symmetric about μ_0 and indeed will depend on μ only through $(\sqrt{n}/\sigma)|\mu-\mu_0|$. This curve with the abscissa changed from μ to $d = (\sqrt{n}/\sigma)|\mu-\mu_0|$ is presented in Figure 8.2 when $\alpha = .05$.

EXAMPLE 8.3c For the problem presented in Example 8.3a, let us determine the probability of accepting the null hypothesis that $\mu = 8$ when the actual value sent is 10. To do so, we compute

$$\frac{\sqrt{n}}{\sigma}(\mu_0-\mu) = -\frac{\sqrt{5}}{2} \times 2 = -\sqrt{5}$$

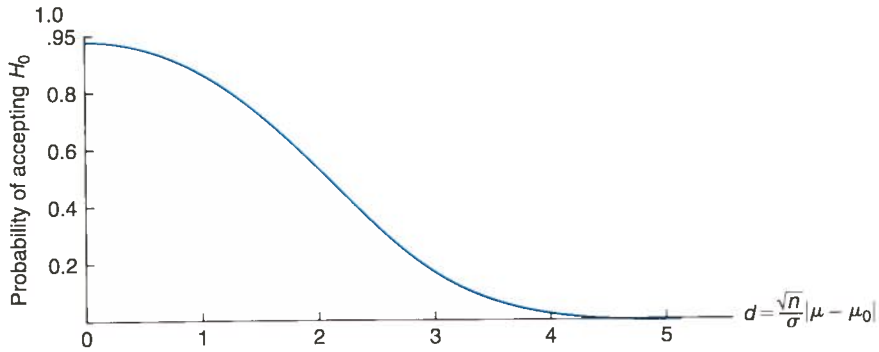


FIGURE 8.2 *The OC curve for the two-sided normal test for significance level* $\alpha = .05$.

As $z_{.025} = 1.96$, the desired probability is, from Equation 8.3.4,

$$\begin{aligned}
&\Phi(-\sqrt{5} + 1.96) - \Phi(-\sqrt{5} - 1.96) \\
&\quad = 1 - \Phi(\sqrt{5} - 1.96) - [1 - \Phi(\sqrt{5} + 1.96)] \\
&\quad = \Phi(4.196) - \Phi(.276) \\
&\quad = .392 \quad ■
\end{aligned}$$

REMARK

The function $1 - \beta(\mu)$ is called the *power-function* of the test. Thus, for a given value μ, the power of the test is equal to the probability of rejection when μ is the true value. ■

The operating characteristic function is useful in determining how large the random sample need be to meet certain specifications concerning type II errors. For instance, suppose that we desire to determine the sample size n necessary to ensure that the probability of accepting $H_0 : \mu = \mu_0$ when the true mean is actually μ_1 is approximately β. That is, we want n to be such that

$$\beta(\mu_1) \approx \beta$$

But from Equation 8.3.4, this is equivalent to

$$\Phi\left(\frac{\sqrt{n}(\mu_0 - \mu_1)}{\sigma} + z_{\alpha/2}\right) - \Phi\left(\frac{\sqrt{n}(\mu_0 - \mu_1)}{\sigma} - z_{\alpha/2}\right) \approx \beta \qquad (8.3.5)$$

Although the foregoing cannot be analytically solved for n, a solution can be obtained by using the standard normal distribution table. In addition, an approximation for n can be derived from Equation 8.3.5 as follows. To start, suppose that $\mu_1 > \mu_0$. Then, because this implies that

$$\frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} - z_{\alpha/2} \leq -z_{\alpha/2}$$

it follows, since Φ is an increasing function, that

$$\Phi\left(\frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} - z_{\alpha/2}\right) \le \Phi(-z_{\alpha/2}) = P\{Z \le -z_{\alpha/2}\} = P\{Z \ge z_{\alpha/2}\} = \alpha/2$$

Hence, we can take

$$\Phi\left(\frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} - z_{\alpha/2}\right) \approx 0$$

and so from Equation 8.3.5

$$\beta \approx \Phi\left(\frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} + z_{\alpha/2}\right) \tag{8.3.6}$$

or, since

$$\beta = P\{Z > z_\beta\} = P\{Z < -z_\beta\} = \Phi(-z_\beta)$$

we obtain from Equation 8.3.6 that

$$-z_\beta \approx (\mu_0 - \mu_1)\frac{\sqrt{n}}{\sigma} + z_{\alpha/2}$$

or

$$n \approx \frac{(z_{\alpha/2} + z_\beta)^2\sigma^2}{(\mu_1 - \mu_0)^2} \tag{8.3.7}$$

In fact, the same approximation would result when $\mu_1 < \mu_0$ (the details are left as an exercise) and so Equation 8.3.7 is in all cases a reasonable approximation to the sample size necessary to ensure that the type II error at the value $\mu = \mu_1$ is approximately equal to β.

EXAMPLE 8.3d For the problem of Example 8.3a, how many signals need be sent so that the .05 level test of $H_0 : \mu = 8$ has at least a 75 percent probability of rejection when $\mu = 9.2$?

SOLUTION Since $z_{.025} = 1.96, z_{.25} = .67$, the approximation 8.3.7 yields

$$n \approx \frac{(1.96 + .67)^2}{(1.2)^2}4 = 19.21$$

Hence a sample of size 20 is needed. From Equation 8.3.4, we see that with $n = 20$

$$\begin{aligned}\beta(9.2) &= \Phi\left(-\frac{1.2\sqrt{20}}{2} + 1.96\right) - \Phi\left(-\frac{1.2\sqrt{20}}{2} - 1.96\right)\\ &= \Phi(-.723) - \Phi(-4.643)\end{aligned}$$

$$\approx 1 - \Phi(.723)$$
$$\approx .235$$

Therefore, if the message is sent 20 times, then there is a 76.5 percent chance that the null hypothesis $\mu = 8$ will be rejected when the true mean is 9.2. ■

8.3.1.1 ONE-SIDED TESTS

In testing the null hypothesis that $\mu = \mu_0$, we have chosen a test that calls for rejection when $\overline{X}$ is far from μ_0. That is, a very small value of $\overline{X}$ or a very large value appears to make it unlikely that μ (which $\overline{X}$ is estimating) could equal μ_0. However, what happens when the only alternative to μ being equal to μ_0 is for μ to be greater than μ_0? That is, what happens when the alternative hypothesis to $H_0 : \mu = \mu_0$ is $H_1 : \mu > \mu_0$? Clearly, in this latter case we would not want to reject H_0 when $\overline{X}$ is small (since a small $\overline{X}$ is more likely when H_0 is true than when H_1 is true). Thus, in testing

$$H_0 : \mu = \mu_0 \qquad \text{versus} \qquad H_1 : \mu > \mu_0 \tag{8.3.8}$$

we should reject H_0 when $\overline{X}$, the point estimate of μ_0, is much greater than μ_0. That is, the critical region should be of the following form:

$$C = \{(X_1, \ldots, X_n) : \overline{X} - \mu_0 > c\}$$

Since the probability of rejection should equal α when H_0 is true (that is, when $\mu = \mu_0$), we require that c be such that

$$P_{\mu_0}\{\overline{X} - \mu_0 > c\} = \alpha \tag{8.3.9}$$

But since

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$$

has a standard normal distribution when H_0 is true, Equation 8.3.9 is equivalent to

$$P\left\{Z > \frac{c\sqrt{n}}{\sigma}\right\} = \alpha$$

when Z is a standard normal. But since

$$P\{Z > z_\alpha\} = \alpha$$

we see that

$$c = \frac{z_\alpha \sigma}{\sqrt{n}}$$

Hence, the test of the hypothesis 8.3.8 is to reject H_0 if $\overline{X} - \mu_0 > z_\alpha \sigma/\sqrt{n}$, and accept otherwise; or, equivalently, to

$$\begin{aligned} &\text{accept} \quad H_0 \quad \text{if} \quad \frac{\sqrt{n}}{\sigma}(\overline{X} - \mu_0) \le z_\alpha \\ &\text{reject} \quad H_0 \quad \text{if} \quad \frac{\sqrt{n}}{\sigma}(\overline{X} - \mu_0) > z_\alpha \end{aligned} \tag{8.3.10}$$

This is called a *one-sided* critical region (since it calls for rejection only when $\overline{X}$ is large). Correspondingly, the hypothesis testing problem

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu > \mu_0$$

is called a one-sided testing problem (in contrast to the *two-sided* problem that results when the alternative hypothesis is $H_1 : \mu \neq \mu_0$).

To compute the p-value in the one-sided test, Equation 8.3.10, we first use the data to determine the value of the statistic $\sqrt{n}(\overline{X} - \mu_0)/\sigma$. The p-value is then equal to the probability that a standard normal would be at least as large as this value.

EXAMPLE 8.3e Suppose in Example 8.3a that we know in advance that the signal value is at least as large as 8. What can be concluded in this case?

SOLUTION To see if the data are consistent with the hypothesis that the mean is 8, we test

$$H_0 : \mu = 8$$

against the one-sided alternative

$$H_1 : \mu > 8$$

The value of the test statistic is $\sqrt{n}(\overline{X} - \mu_0)/\sigma = \sqrt{5}(9.5 - 8)/2 = 1.68$, and the p-value is the probability that a standard normal would exceed 1.68, namely,

$$p\text{-value} = 1 - \Phi(1.68) = .0465$$

Since the test would call for rejection at all significance levels greater than or equal to .0465, it would, for instance, reject the null hypothesis at the $\alpha = .05$ level of significance. ■

The operating characteristic function of the one-sided test, Equation 8.3.10,

$$\beta(\mu) = P_\mu\{\text{accepting } H_0\}$$

can be obtained as follows:

$$\begin{aligned}\beta(\mu) &= P_\mu\left\{\overline{X} \le \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}\right\} \\ &= P\left\{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \le \frac{\mu_0-\mu}{\sigma/\sqrt{n}} + z_\alpha\right\} \\ &= P\left\{Z \le \frac{\mu_0-\mu}{\sigma/\sqrt{n}} + z_\alpha\right\}, \quad Z \sim \mathcal{N}(0,1)\end{aligned}$$

where the last equation follows since $\sqrt{n}(\overline{X}-\mu)/\sigma$ has a standard normal distribution. Hence we can write

$$\beta(\mu) = \Phi\left(\frac{\mu_0-\mu}{\sigma/\sqrt{n}} + z_\alpha\right)$$

Since Φ, being a distribution function, is increasing in its argument, it follows that $\beta(\mu)$ decreases in μ, which is intuitively pleasing since it certainly seems reasonable that the larger the true mean μ, the less likely it should be to conclude that $\mu \le \mu_0$. Also since $\Phi(z_\alpha) = 1-\alpha$, it follows that

$$\beta(\mu_0) = 1-\alpha$$

The test given by Equation 8.3.10, which was designed to test $H_0 : \mu = \mu_0$ versus $H_1 : \mu > \mu_0$, can also be used to test, at level of significance α, the one-sided hypothesis

$$H_0 : \mu \le \mu_0$$

versus

$$H_1 : \mu > \mu_0$$

To verify that it remains a level α test, we need to show that the probability of rejection is never greater than α when H_0 is true. That is, we must verify that

$$1-\beta(\mu) \le \alpha \qquad \text{for all } \mu \le \mu_0$$

or

$$\beta(\mu) \ge 1-\alpha \qquad \text{for all } \mu \le \mu_0$$

But it has previously been shown that for the test given by Equation 8.3.10, $\beta(\mu)$ decreases in μ and $\beta(\mu_0) = 1-\alpha$. This gives that

$$\beta(\mu) \ge \beta(\mu_0) = 1-\alpha \qquad \text{for all } \mu \le \mu_0$$

which shows that the test given by Equation 8.3.10 remains a level α test for $H_0 : \mu \le \mu_0$ against the alternative hypothesis $H_1 : \mu \le \mu_0$.

REMARK

We can also test the one-sided hypothesis

$$H_0 : \mu = \mu_0 \quad (\text{or } \mu \geq \mu_0) \qquad \text{versus} \quad H_1 : \mu < \mu_0$$

at significance level α by

$$\begin{aligned} &\text{accepting} \quad H_0 \quad \text{if} \quad \frac{\sqrt{n}}{\sigma}(\overline{X} - \mu_0) \geq -z_\alpha \\ &\text{rejecting} \quad H_0 \quad \text{otherwise} \end{aligned}$$

This test can alternatively be performed by first computing the value of the test statistic $\sqrt{n}(\overline{X} - \mu_0)/\sigma$. The p-value would then equal the probability that a standard normal would be less than this value, and the hypothesis would be rejected at any significance level greater than or equal to this p-value.

EXAMPLE 8.3f All cigarettes presently on the market have an average nicotine content of at least 1.6 mg per cigarette. A firm that produces cigarettes claims that it has discovered a new way to cure tobacco leaves that will result in the average nicotine content of a cigarette being less than 1.6 mg. To test this claim, a sample of 20 of the firm's cigarettes were analyzed. If it is known that the standard deviation of a cigarette's nicotine content is .8 mg, what conclusions can be drawn, at the 5 percent level of significance, if the average nicotine content of the 20 cigarettes is 1.54?

Note: The above raises the question of how we would know in advance that the standard deviation is .8. One possibility is that the variation in a cigarette's nicotine content is due to variability in the amount of tobacco in each cigarette and not on the method of curing that is used. Hence, the standard deviation can be known from previous experience.

SOLUTION We must first decide on the appropriate null hypothesis. As was previously noted, our approach to testing is not symmetric with respect to the null and the alternative hypotheses since we consider only tests having the property that their probability of rejecting the null hypothesis when it is true will never exceed the significance level α. Thus, whereas rejection of the null hypothesis is a strong statement about the data not being consistent with this hypothesis, an analogous statement cannot be made when the null hypothesis is accepted. Hence, since in the preceding example we would like to endorse the producer's claims only when there is substantial evidence for it, we should take this claim as the alternative hypothesis. That is, we should test

$$H_0 : \mu \geq 1.6 \qquad \text{versus} \qquad H_1 : \mu < 1.6$$

Now, the value of the test statistic is

$$\sqrt{n}(\overline{X} - \mu_0)/\sigma = \sqrt{20}(1.54 - 1.6)/.8 = -.336$$

and so the p-value is given by

$$
\begin{aligned}
p\text{-value} &= P\{Z < -.336\}, \qquad Z \sim N(0,1) \\
&= .368
\end{aligned}
$$

Since this value is greater than .05, the foregoing data do not enable us to reject, at the .05 percent level of significance, the hypothesis that the mean nicotine content exceeds 1.6 mg. In other words, the evidence, although supporting the cigarette producer's claim, is not strong enough to prove that claim. ■

REMARKS

(a) There is a direct analogy between confidence interval estimation and hypothesis testing. For instance, for a normal population having mean μ and known variance σ^2, we have shown in Section 7.3 that a $100(1-\alpha)$ percent confidence interval for μ is given by

$$\mu \in \left(\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

where $\overline{x}$ is the observed sample mean. More formally, the preceding confidence interval statement is equivalent to

$$P\left\{\mu \in \left(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)\right\} = 1 - \alpha$$

Hence, if $\mu = \mu_0$, then the probability that μ_0 will fall in the interval

$$\left(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

is $1 - \alpha$, implying that a significance level α test of $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$ is to reject H_0 when

$$\mu_0 \notin \left(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

Similarly, since a $100(1-\alpha)$ percent one-sided confidence interval for μ is given by

$$\mu \in \left(\overline{X} - z_{\alpha}\frac{\sigma}{\sqrt{n}}, \infty\right)$$

it follows that an α-level significance test of $H_0 : \mu \leq \mu_0$ versus $H_1 : \mu > \mu_0$ is to reject H_0 when $\mu_0 \notin (\overline{X} - z_\alpha\sigma/\sqrt{n}, \infty)$ — that is, when $\mu_0 < \overline{X} - z_\alpha\sigma/\sqrt{n}$.

TABLE 8.1 $X_1, \ldots, X_n$ *Is a Sample from a* $\mathcal{N}(\mu, \sigma^2)$ *Population* σ^2 *Is Known,* $\overline{X} = \sum_{i=1}^{n} X_i/n$

H_0	H_1	Test Statistic TS	Significance Level α Test	p-Value if $TS = t$
$\mu = \mu_0$	$\mu \neq \mu_0$	$\sqrt{n}(\overline{X} - \mu_0)/\sigma$	Reject if $\lvert TS \rvert > z_{\alpha/2}$	$2P\{Z \geq \lvert t \rvert\}$
$\mu \leq \mu_0$	$\mu > \mu_0$	$\sqrt{n}(\overline{X} - \mu_0)/\sigma$	Reject if $TS > z_\alpha$	$P\{Z \geq t\}$
$\mu \geq \mu_0$	$\mu < \mu_0$	$\sqrt{n}(\overline{X} - \mu_0)/\sigma$	Reject if $TS < -z_\alpha$	$P\{Z \leq t\}$

Z is a standard normal random variable.

(b) A Remark on Robustness A test that performs well even when the underlying assumptions on which it is based are violated is said to be ***robust***. For instance, the tests of Sections 8.3.1 and 8.3.1.1 were derived under the assumption that the underlying population distribution is normal with known variance σ^2. However, in deriving these tests, this assumption was used only to conclude that $\overline{X}$ also has a normal distribution. But, by the central limit theorem, it follows that for a reasonably large sample size, $\overline{X}$ will approximately have a normal distribution no matter what the underlying distribution. Thus we can conclude that these tests will be relatively robust for any population distribution with variance σ^2.

Table 8.1 summarizes the tests of this subsection.

8.3.2 Case of Unknown Variance: The t-Test

Up to now we have supposed that the only unknown parameter of the normal population distribution is its mean. However, the more common situation is one where the mean μ and variance σ^2 are both unknown. Let us suppose this to be the case and again consider a test of the hypothesis that the mean is equal to some specified value μ_0. That is, consider a test of

$$H_0 : \mu = \mu_0$$

versus the alternative

$$H_1 : \mu \neq \mu_0$$

It should be noted that the null hypothesis is not a simple hypothesis since it does not specify the value of σ^2.

As before, it seems reasonable to reject H_0 when the sample mean $\overline{X}$ is far from μ_0. However, how far away it need be to justify rejection will depend on the variance σ^2. Recall that when the value of σ^2 was known, the test called for rejecting H_0 when $|\overline{X} - \mu_0|$ exceeded $z_{\alpha/2}\sigma/\sqrt{n}$ or, equivalently, when

$$\left| \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \right| > z_{\alpha/2}$$

Now when σ^2 is no longer known, it seems reasonable to estimate it by

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$$

and then to reject H_0 when

$$\left|\frac{\overline{X} - \mu_0}{S/\sqrt{n}}\right|$$

is large.

To determine how large a value of the statistic

$$\left|\frac{\sqrt{n}(\overline{X} - \mu_0)}{S}\right|$$

to require for rejection, in order that the resulting test have significance level α, we must determine the probability distribution of this statistic when H_0 is true. However, as shown in Section 6.5, the statistic T, defined by

$$T = \frac{\sqrt{n}(\overline{X} - \mu_0)}{S}$$

has, when $\mu = \mu_0$, a t-distribution with $n - 1$ degrees of freedom. Hence,

$$P_{\mu_0}\left\{-t_{\alpha/2,\,n-1} \le \frac{\sqrt{n}(\overline{X} - \mu_0)}{S} \le t_{\alpha/2,\,n-1}\right\} = 1 - \alpha \tag{8.3.11}$$

where $t_{\alpha/2,n-1}$ is the $100\,\alpha/2$ upper percentile value of the t-distribution with $n-1$ degrees of freedom. (That is, $P\{T_{n-1} \ge t_{\alpha/2,n-1}\} = P\{T_{n-1} \le -t_{\alpha/2,n-1}\} = \alpha/2$ when T_{n-1} has a t-distribution with $n - 1$ degrees of freedom.) From Equation 8.3.11 we see that the appropriate significance level α test of

$$H_0 : \mu = \mu_0 \qquad \text{versus} \qquad H_1 : \mu \ne \mu_0$$

is, when σ^2 is unknown, to

$$\begin{aligned} &\text{accept } H_0 \quad \text{if} \quad \left|\frac{\sqrt{n}(\overline{X} - \mu_0)}{S}\right| \le t_{\alpha/2,\,n-1} \\ &\text{reject } H_0 \quad \text{if} \quad \left|\frac{\sqrt{n}(\overline{X} - \mu_0)}{S}\right| > t_{\alpha/2,\,n-1} \end{aligned} \tag{8.3.12}$$

FIGURE 8.3 *The two-sided t-test.*

The test defined by Equation 8.3.12 is called a *two-sided t-test.* It is pictorially illustrated in Figure 8.3.

If we let t denote the observed value of the test statistic $T = \sqrt{n}(\overline{X} - \mu_0)/S$, then the p-value of the test is the probability that $|T|$ would exceed $|t|$ when H_0 is true. That is, the p-value is the probability that the absolute value of a t-random variable with $n-1$ degrees of freedom would exceed $|t|$. The test then calls for rejection at all significance levels higher than the p-value and acceptance at all lower significance levels.

Program 8.3.2 computes the value of the test statistic and the corresponding p-value. It can be applied both for one- and two-sided tests. (The one-sided material will be presented shortly.)

EXAMPLE 8.3g Among a clinic's patients having blood cholesterol levels ranging in the medium to high range (at least 220 milliliters per deciliter of serum), volunteers were recruited to test a new drug designed to reduce blood cholesterol. A group of 50 volunteers was given the drug for 1 month and the changes in their blood cholesterol levels were noted. If the average change was a reduction of 14.8 with a sample standard deviation of 6.4, what conclusions can be drawn?

SOLUTION Let us start by testing the hypothesis that the change could be due solely to chance — that is, that the 50 changes constitute a normal sample with mean 0. Because the value of the t-statistic used to test the hypothesis that a normal mean is equal to 0 is

$$T = \sqrt{n}\,\overline{X}/S = \sqrt{50}\,14.8/6.4 = 16.352$$

it is clear that we should reject the hypothesis that the changes were solely due to chance. Unfortunately, however, we are not justified at this point in concluding that the changes were due to the specific drug used and not to some other possibility. For instance, it is well known that any medication received by a patient (whether or not this medication is directly relevant to the patient's suffering) often leads to an improvement in the patient's condition — the so-called placebo effect. Also, another possibility that may need to be taken into account would be the weather conditions during the month of testing, for it is certainly conceivable that this affects blood cholesterol level. Indeed, it must be concluded that the foregoing was a very poorly designed experiment, for in order to test whether a specific treatment has an effect on a disease that may be affected by many things, we should try to design the experiment so as to neutralize all other possible causes. The accepted approach for accomplishing this is to divide the volunteers at random into two groups — one group to receive the drug and the other to receive a placebo (that is, a tablet that looks and tastes like the actual drug but has no physiological effect). The volunteers should not be told whether they are in the actual or control group, and indeed it is best if even the clinicians do not have this information (the so-called double-blind test) so as not to allow their own biases to play a role. Since the two groups are chosen at random from among the volunteers, we can now hope that on average all factors affecting the two groups will be the same except that one received the actual drug and the other a placebo. Hence, any difference in performance between the groups can be attributed to the drug. ■

EXAMPLE 8.3h A public health official claims that the mean home water use is 350 gallons a day. To verify this claim, a study of 20 randomly selected homes was instigated with the result that the average daily water uses of these 20 homes were as follows:

340	344	362	375
356	386	354	364
332	402	340	355
362	322	372	324
318	360	338	370

Do the data contradict the official's claim?

SOLUTION To determine if the data contradict the official's claim, we need to test

$$H_0 : \mu = 350 \qquad \text{versus} \qquad H_1 : \mu \neq 350$$

This can be accomplished by running Program 8.3.2 or, if it is incovenient to utilize, by noting first that the sample mean and sample standard deviation of the preceding data set are

$$\overline{X} = 353.8, \qquad S = 21.8478$$

Thus, the value of the test statistic is

$$T = \frac{\sqrt{20}(3.8)}{21.8478} = .7778$$

Because this is less than $t_{.05,19} = 1.730$, the null hypothesis is accepted at the 10 percent level of significance. Indeed, the p-value of the test data is

$$p\text{-value} = P\{|T_{19}| > .7778\} = 2P\{T_{19} > .7778\} = .4462$$

indicating that the null hypothesis would be accepted at any reasonable significance level, and thus that the data are not inconsistent with the claim of the health official. ■

We can use a one-sided t-test to test the hypothesis

$$H_0 : \mu = \mu_0 \qquad (\text{or } H_0 : \mu \leq \mu_0)$$

against the one-sided alternative

$$H_1 : \mu > \mu_0$$

The significance level α test is to

$$\begin{aligned} &\text{accept} \quad H_0 \quad \text{if} \quad \frac{\sqrt{n}(\overline{X} - \mu_0)}{S} \leq t_{\alpha, n-1} \\ &\text{reject} \quad H_0 \quad \text{if} \quad \frac{\sqrt{n}(\overline{X} - \mu_0)}{S} > t_{\alpha, n-1} \end{aligned} \tag{8.3.13}$$

If $\sqrt{n}(\overline{X} - \mu_0)/S = v$, then the p-value of the test is the probability that a t-random variable with $n - 1$ degrees of freedom would be at least as large as v.

The significance level α test of

$$H_0 : \mu = \mu_0 \qquad (\text{or } H_0 : \mu \geq \mu_0)$$

versus the alternative

$$H_1 : \mu < \mu_0$$

is to

$$\begin{aligned} &\text{accept} \quad H_0 \quad \text{if} \quad \frac{\sqrt{n}(\overline{X} - \mu_0)}{S} \geq -t_{\alpha, n-1} \\ &\text{reject} \quad H_0 \quad \text{if} \quad \frac{\sqrt{n}(\overline{X} - \mu_0)}{S} < -t_{\alpha, n-1} \end{aligned}$$

The p-value of this test is the probability that a t-random variable with $n - 1$ degrees of freedom would be less than or equal to the observed value of $\sqrt{n}(\overline{X} - \mu_0)/S$.

EXAMPLE 8.3i The manufacturer of a new fiberglass tire claims that its average life will be at least 40,000 miles. To verify this claim a sample of 12 tires is tested, with their lifetimes (in 1,000s of miles) being as follows:

Tire	1	2	3	4	5	6	7	8	9	10	11	12
Life	36.1	40.2	33.8	38.5	42	35.8	37	41	36.8	37.2	33	36

Test the manufacturer's claim at the 5 percent level of significance.

SOLUTION To determine whether the foregoing data are consistent with the hypothesis that the mean life is at least 40,000 miles, we will test

$$H_0 : \mu \geq 40{,}000 \qquad \text{versus} \qquad H_1 : \mu < 40{,}000$$

A computation gives that

$$\overline{X} = 37.2833, \qquad S = 2.7319$$

and so the value of the test statistic is

$$T = \frac{\sqrt{12}(37.2833 - 40)}{2.7319} = -3.4448$$

Since this is less than $-t_{.05,11} = -1.796$, the null hypothesis is rejected at the 5 percent level of significance. Indeed, the p-value of the test data is

$$p\text{-value} = P\{T_{11} < -3.4448\} = P\{T_{11} > 3.4448\} = .0028$$

indicating that the manufacturer's claim would be rejected at any significance level greater than .003. ■

The preceding could also have been obtained by using Program 8.3.2, as illustrated in Figure 8.4.

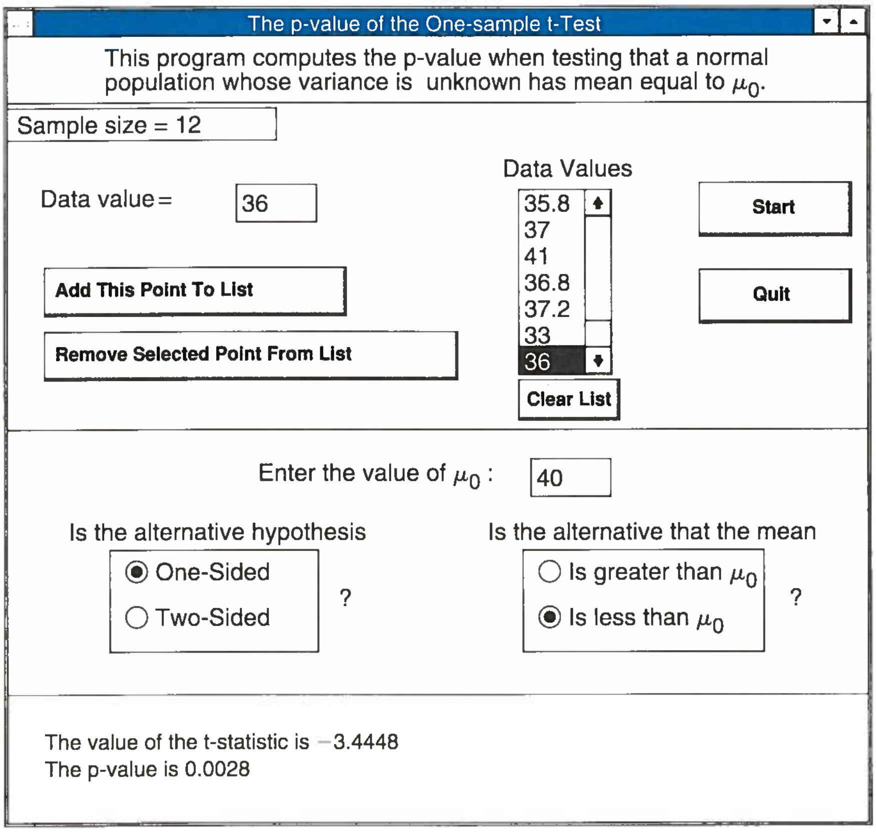


FIGURE 8.4